Лабораторная работа 1: Применение алгоритмов кластеризации для диагностики кризиса теплообмена в ЯЭУ.

Установка необходимых библиотек и инструментов

Jupyter Notebook - это веб-приложение с открытым исходным кодом, которое вы можете использовать для создания и обмена документами, которые содержат живой код, уравнения, визуализацию, текст и научные исследования.

Jupyter Notebooks - это побочный проект проекта IPython, который раньше имел сам проект IPython Notebook. Название Jupyter происходит от основных поддерживаемых языков программирования, которые он поддерживает: Julia, Python и R. Jupyter поставляется с ядром IPython, которое позволяет писать свои программы на Python, но в настоящее время существует более 100 других ядер, которые вы можете использовать. также можно использовать.

pip install jupyter – установка Jupyter Notebook.

jupyter notebook – запуск Jupyter Notebook.

Numpy - это библиотека Python для вычислительно эффективных операций с многомерными массивами, предназначенная в основном для научных вычислений.

import numpy as np

Pandas - это библиотека Python, предоставляющая широкие возможности для анализа данных. С ее помощью очень удобно загружать, обрабатывать и анализировать табличные данные с помощью SQL-подобных запросов.

import pandas as pd

Основными структурами данных в Pandas являются классы Series и DataFrame. Первый из них представляет собой одномерный индексированный массив данных некоторого фиксированного типа. Второй - это двумерная структура данных, представляющая собой таблицу, каждый столбец которой содержит данные одного типа. Можно представлять её как словарь объектов типа Series.

Выполнение лабораторной работы

1. Прочтите данные из файлов varX.csv, targetX.csv (где X – номер варианта).

Функции, которые могут пригодиться при решении: pd.read_csv()

2. Транспонируйте исходную матрицу. Каждый столбец будет спектром с 200 частотами.
3. Отобразите несколько первых и несколько последних записей.

Функции, которые могут пригодиться при решении: .head(), .tail().

4. Постройте графики временных реализаций каждого спектра с помощью цикла или встроенных средств библиотеки pandas для визуализации. Используйте функцию plot() из библиотеки matplotlib.
5. Постройте временные реализации каждого спектра на одном графике. Используйте функцию plot() из библиотеки matplotlib..
6. Найдите мощность всего спектра по формуле

$$P = \sum_{i-1}^{200} S(f_i)$$

7. Выведите распределение целевой переменной targetX.csv. Зафиксируйте индекс начала кризиса теплообмена (значения равные 2).
8. Постройте график мощности спектра и начертите на нем вертикальную линию, которая будет разъединять участок без кризиса и с кризисом.

9. Выведите описательные статистики данных мощности спектра до кризиса (без кризиса) и с кризисом теплообмена.

Функции, которые могут пригодиться при решении: .describe()

10. Отдельно выведите в рабочую область средние значения мощности спектра до кризиса и после.
11. Постройте диаграммы ящиков с усами (boxplots) мощности спектра до кризиса и после на одном графики и сравните их. Опишите все наблюдения по построенным диаграммам.
12. Найдите среднюю частоту спектра, используя следующую формулу:

$$\overline{f} = \frac{\sum_{i=1}^{200} S(f_i) * f_i}{\sum_{i=1}^{200} S(f_i)}$$

13. Постройте график значений средних частот спектра и начертите на нем вертикальную линию, которая будет разъединять участок без кризиса и с кризисом.
14. Выведите описательные статистики данных средних значений спектра до кризиса (без кризиса) и с кризисом теплообмена.

Функции, которые могут пригодиться при решении: .describe()

15. Отдельно выведите в рабочую область средние значения средних частот спектра до кризиса и после.
16. Постройте диаграммы ящиков с усами (boxplots) средних частот спектра до кризиса и после на одном графики и сравните их. Опишите все наблюдения по построенным диаграммам. Используется функцию boxplots() из библиотеки matplotlib.
17. Постройте график, где по оси X будет отложена мощность, а по У отложена средняя частота. Раскрасьте точки на графике с помощью значений вектора целевой переменной.
18. Постройте 5 графиков с 5 парами (т.е. на каждом графике по 2) случайных частот, выбранных из исходного набора данных.

Раскрасьте точки на графике с помощью значений вектора целевой переменной. Проведите прямую, равноудаленную от точек каждого класса (можно использовать МНК).

19. Примените метод понижения размерности (метод главных компонент) к исходному набору данных с частотами спектров. Визуализируйте 2 первые главные компоненты на плоскости и раскрасьте точки на графике с помощью значений вектора целевой переменной. (можно воспользоваться следующей библиотекой *from sklearn.decomposition import PCA)*. Не забудьте выполнить масштабирование многомерных данных перед их визуализацией на плоскости и понижением размерности.
20. Примените не менее 3-ех методов кластеризации (*KMeans*, *SpectralClustering*, *AgglomerativeClustering*, *DBSCA*N и др.) к исходным данным спектров и выполните их кластеризацию на 2 класса (без кризиса и с кризисом). Проверьте качество кластеризации по метрике homogeneity_completeness_v_measure из библиотеки sklearn, модуля metrics (*from sklearn.metrics import homogeneity_completeness_v_measure)*. Алгоритмы кластеризации можно найти в модуле *sklearn.clustering*.
21. Повысьте точность работы алгоритмов с помощью выбора информативных признаков (частот или спектров) из исходного набора данных. Заново примените алгоритмы и добейтесь наилучшей точности работы алгоритмов.
22. Оформите отчет по лабораторной работе в формате .ipynb с заголовками, комментариями, рисунками (с заголовками и названиями осей), ответами на контрольные вопросы, а также выводами о проделанной работе. Перед первым заголовком должно быть ваше ФИО и название группы. Назовите файл ФИО_lab1.ipynb и сделайте файл .pdf с таким же названием, а затем сдайте оба файла преподавателю.

Контрольные вопросы

1. Какие существуют алгоритмы кластерного анализа данных? Назовите не менее 3-ех и опишите их суть с математической точки зрения и расскажите чем они отличаются друг от друга.
2. Какие метрики используются для оценки качества работы алгоритма кластеризации данных? Опишите данные метрики с математической точки зрения и скажите чем они отличаются друг от друга.
3. Каким способом можно повысить качество работы алгоритмов кластеризации?

Лабораторная работа 2: Использование метода ближайших соседей (KNN) для решения задачи классификации. Настройка гиперпараметров модели машинного обучения.

Установка необходимых библиотек и инструментов

Jupyter Notebook - это веб-приложение с открытым исходным кодом, которое вы можете использовать для создания и обмена документами, которые содержат живой код, уравнения, визуализацию, текст и научные исследования.

Jupyter Notebooks - это побочный проект проекта IPython, который раньше имел сам проект IPython Notebook. Название Jupyter происходит от основных поддерживаемых языков программирования, которые он поддерживает: Julia, Python и R. Jupyter поставляется с ядром IPython, которое позволяет писать свои программы на Python, но в настоящее время существует более 100 других ядер, которые вы можете использовать. также можно использовать.

pip install jupyter – установка Jupyter Notebook.

jupyter notebook – запуск Jupyter Notebook.

Numpy - это библиотека Python для вычислительно эффективных операций с многомерными массивами, предназначенная в основном для научных вычислений.

import numpy as np

Pandas - это библиотека Python, предоставляющая широкие возможности для анализа данных. С ее помощью очень удобно загружать, обрабатывать и анализировать табличные данные с помощью SQL-подобных запросов.

import pandas as pd

Основными структурами данных в Pandas являются классы Series и DataFrame. Первый из них представляет собой одномерный индексированный массив данных некоторого фиксированного типа. Второй - это двумерная структура данных, представляющая собой таблицу, каждый столбец которой содержит данные одного типа. Можно представлять её как словарь объектов типа Series.

Выполнение лабораторной работы

1. Прочтите данные из файлов varX.csv, targetX.csv (где X – номер варианта).

Функции, которые могут пригодиться при решении: pd.read_csv()

2. Транспонируйте исходную матрицу. Каждый столбец будет спектром с 200 частотами.
3. Отобразите несколько первых и несколько последних записей.

Функции, которые могут пригодиться при решении: .head(), .tail().

4. Разбейте данные на обучающую и проверочную выборки в пропорции 70 на 30 с помощью функции *train_test_split()* из библиотеки *sklearn*.
5. Примените алгоритм K-ближайших соседей (KNN) к массиву обучающей выборки исходных данных с параметрами, установленными по-умолчанию. Алгоритм KNN можно загрузить используя следующее выражение: *from sklearn.neighbors import KNeighborsClassifier*.
6. Оцените качество модели с помощью метрики *accuracy* и classification report из библиотеки *sklearn* модуля *metrics*.
7. Выполните отбор информативных частот, а затем снова обучите модель *KNN* и оцените качество ее работы на проверочной выборке.
8. Выполните подбор гиперпараметров модели *KNN* с помощью *GridSearchCV() (для загрузки класса используйте: from*

sklearn.model_selection import GridSearchCV) из библиотеки sklearn с параметром кросс-валидации cv = 5. Основным настраиваемым гиперпараметром алгоритма является *n_neighbors*.

9. Заново обучите модель с подобранными гиперпараметрами на обучающей выборке и оцените качество ее работы на проверочной.
10. Оформите отчет по лабораторной работе в формате ipynb с заголовками, комментариями, рисунками (с заголовками и названиями осей), ответами на контрольные вопросы, а также выводами о проделанной работе. Перед первым заголовком должно быть ваше ФИО и название группы. Назовите файл ФИО_lab2.ipynb и сделайте файл .pdf с таким же названием, а затем сдайте оба файла преподавателю.

Контрольные вопросы

1. Опишите этапы реализации алгоритма KNN и для решения каких задач его можно использовать?
2. Какие метрики используются для оценки качества работы алгоритмов классификации? Опишите данные метрики с математической точки зрения и скажите, чем они отличаются друг от друга.
3. Каким способом можно повысить качество работы алгоритмов классификации?

Лабораторная работа 3: Решающие деревья для задач классификации и регрессии

Описание исходных данных

Данные в формате .csv содержат информацию о средних значениях концентрации четырех измеренных в трансформаторном масле газов (H2; CO; C2H4; C2H2) в различных трансформаторах. Целевые переменные расположены в столбцах *label* (для задачи классификации) и *predicted* (для задачи регрессии).Необходимо по имеющимся данным предсказать тип дефекта трансформатора (частичный разряд, разряд низкой энергии, низкотемпературный перегрев) и прогноз технического состояния и срока службы, а также через какое время концентрации газов достигнут уставки.

Выполнение лабораторной работы

1. Прочтите данные из файлов transformators.csv (набор данных по трансформаторам для классификации), transformators_regression.csv (набор данных по трансформаторам для регрессии).

Функции, которые могут пригодиться при решении: pd.read_csv()

2. Отобразите несколько первых и несколько последних записей.

Функции, которые могут пригодиться при решении: .head(), .tail().

3. Разбейте данные для классификации (transformators.csv) на обучающую и проверочную выборки в пропорции 70 на 30 с помощью функции *train_test_split()* из библиотеки *sklearn*.
4. Примените алгоритм дерева решений (*DecisionTreeClassifier)* к массиву обучающей выборки с параметрами, установленными по-умолчанию. Алгоритм дерева решений для задачи классификации можно загрузить используя следующее выражение: *from sklearn.tree import DecisionTreeClassifier*.

5. Сделайте предсказание на тестовой выборке. Оцените качество модели с помощью метрики *accuracy* и classification report из библиотеки *sklearn* модуля *metrics*.
6. Выполните подбор гиперпараметров модели деревьев классификации с помощью *GridSearchCV() (from sklearn.model_selection import GridSearchCV)* с параметром кросс-валидации cv = 5. Основными настраиваемыми гиперпараметрами алгоритма являются: *max_depth, min_samples_split, min_samples_leaf*. Подробнее о каждом параметре можно прочитать в помощи по модели DTC.
7. Заново обучите модель с подобранными гиперпараметрами на обучающей выборке и оцените качество ее работы на тестовой.
8. Постройте итоговое дерево классификации используя модуль graphviz.
9. Разбейте данные для регрессии (transformators_regression.csv) на обучающую и проверочную выборки в пропорции 70 на 30 с помощью функции *train_test_split()* из библиотеки *sklearn*.
10. Примените алгоритм дерева регрессии (*DecisionTreeRegressor)* массиву обучающей выборки с параметрами, установленными по-умолчанию. Алгоритм дерева решений для задачи регрессии можно загрузить используя следующее выражение: *from sklearn.tree import DecisionTreeRegressor*.
11. Сделайте предсказание на тестовой выборке. Оцените качество модели с помощью метрики *R-square (коэффициент детерминации) и MAE (средняя абсолютная ошибка)* из библиотеки *sklearn* модуля *metrics. (from sklearn.metrics import r2, mean_absolute_error)*
12. Выполните подбор гиперпараметров модели деревьев регрессии с помощью *GridSearchCV() (from sklearn.model_selection import GridSearchCV)* с параметром кросс-валидации cv = 5. Основными настраиваемыми гиперпараметрами алгоритма являются: *max_depth,*

min_samples_split, min_samples_leaf. Подробнее о каждом параметре можно прочитать в помощи по модели DTC.

13. Заново обучите модель с подобранными гиперпараметрами на обучающей выборке и оцените качество ее работы на тестовой (метрики – MAE, R-Square).
14. Постройте итоговое дерево регрессии используя модуль graphviz.
15. Оформите отчет по лабораторной работе в формате ipynb с заголовками, комментариями, рисунками (с заголовками и названиями осей), ответами на контрольные вопросы, а также выводами о проделанной работе. Перед первым заголовком должно быть ваше ФИО и название группы. Назовите файл ФИО_lab3.ipynb и сделайте файл .pdf с таким же названием, а затем сдайте оба файла преподавателю.

Контрольные вопросы

1. Опишите этапы построения алгоритма дерева решений для задачи классификации и регрессии. Чем они отличаются и чем схожи?
2. Какие метрики используются для оценки качества работы алгоритмов при решении задачи регрессии? Опишите данные метрики с математической точки зрения и скажите, чем они отличаются друг от друга.
3. Каким способом можно повысить качество работы алгоритмов регрессии?

Лабораторная работа 4: Применение линейных методов классификации для диагностики кризиса теплообмена в ЯЭУ.

Выполнение лабораторной работы

1. Прочтите данные из файлов varX.csv, targetX.csv (где X – номер варианта).

Функции, которые могут пригодиться при решении: pd.read_csv()

2. Транспонируйте исходную матрицу. Каждый столбец будет спектром с 200 частотами.
3. Отобразите несколько первых и несколько последних записей.

Функции, которые могут пригодиться при решении: .head(), .tail().

4. Выполните отбор информативных частот.
5. Разбейте данные, отобранные на шаге 3, на обучающую и проверочную выборки в пропорции 70 на 30 с помощью функции *train_test_split()* из библиотеки *sklearn*.
6. Последовательно обучите алгоритм логистической регрессии (LogisticRegression), стохастического градиентного спуска (SGDClassifier), классификатор гребневой регрессии (RidgeClassifier) и классификатор лассо (LassoClassifier) на массиве обучающей выборки с параметрами, установленными по-умолчанию. Перечисленные алгоритмы можно загрузить используя модуль библиотеки sklearn – linear_model (например, from sklearn.linear_model import RidgeClassifier)
7. Оцените качество модели с помощью метрики *accuracy* и *classification report* из библиотеки *sklearn* модуля *metrics*. Выберите наилучший алгоритм. Аргументируйте свой выбор.
8. Выполните подбор гиперпараметров для лучшей модели с помощью *GridSearchCV() (для загрузки класса используйте: from sklearn.model_selection import GridSearchCV)* из библиотеки sklearn с

параметром кросс-валидации cv = 5. Подумайте какие параметры стоит настроить. Аргументируйте свой выбор.

9. Заново обучите модель с подобранными гиперпараметрами на обучающей выборке и оцените качество ее работы на проверочной, используя метрики *accuracy* и *classification report* из библиотеки *sklearn* модуля *metrics*.
10. Оформите отчет по лабораторной работе в формате ipynb с заголовками, комментариями, рисунками (с заголовками и названиями осей), ответами на контрольные вопросы, а также выводами о проделанной работе. Перед первым заголовком должно быть ваше ФИО и название группы. Назовите файл ФИО_lab4.ipynb и сделайте файл .pdf с таким же названием, а затем сдайте оба файла преподавателю.

Контрольные вопросы

1. Опишите этапы построения линейных классификаторов. Чем они отличаются и чем схожи?
2. Что означает L-1 и L-2 регуляризация?
3. В чем заключается метод стохастического градиентного спуска? Где и когда его можно использовать?

Лабораторная работа 5: Мультиклассификация с помощью SVM моделей (трансформаторы)

Описание исходных данных

Данные в формате .csv содержат информацию о средних значениях концентрации четырех измеренных в трансформаторном масле газов (H2; CO; C2H4; C2H2) в различных трансформаторах. Целевые переменные расположены в столбцах *label* (для задачи классификации) и *predicted* (для задачи регрессии).Необходимо по имеющимся данным предсказать тип дефекта трансформатора (частичный разряд, разряд низкой энергии, низкотемпературный перегрев) и прогноз технического состояния и срока службы, а также через какое время концентрации газов достигнут уставки.

Выполнение лабораторной работы

1. Прочтите данные из файлов transformators.csv (набор данных по трансформаторам для классификации).

Функции, которые могут пригодиться при решении: pd.read_csv()

2. Отобразите несколько первых и несколько последних записей.

Функции, которые могут пригодиться при решении: .head(), .tail().

3. Разбейте данные для классификации (transformators.csv) на обучающую и проверочную выборки в пропорции 70 на 30 с помощью функции *train_test_split()* из библиотеки *sklearn*.
4. Примените метод опорных векторов для классификации (*SVC)* к массиву обучающей выборки с параметрами, установленными по-умолчанию. Метод опорных векторов задачи классификации можно загрузить используя следующее выражение: *from sklearn.svm import SVC*.

5. Сделайте предсказание на тестовой выборке. Оцените качество модели с помощью метрики *accuracy* и *classification report* из библиотеки *sklearn* модуля *metrics*.
6. Выполните подбор гиперпараметров модели *SVC* с помощью *GridSearchCV() (from sklearn.model_selection import GridSearchCV)* с параметром кросс-валидации *cv* = 5. Подумайте какие параметры необходимо настроить. Аргументируйте свой выбор.
7. Заново обучите модель с подобранными гиперпараметрами на обучающей выборке и оцените качество ее работы на тестовой.
8. Оформите отчет по лабораторной работе в формате ipynb с заголовками, комментариями, рисунками (с заголовками и названиями осей), ответами на контрольные вопросы, а также выводами о проделанной работе. Перед первым заголовком должно быть ваше ФИО и название группы. Назовите файл ФИО_lab5.ipynb и сделайте файл .pdf с таким же названием, а затем сдайте оба файла преподавателю.

Контрольные вопросы

1. Опишите этапы построения алгоритма классификации SVC. В чем его отличие от других линейных алгоритмов?
2. Какие гиперпараметры влияют на качество работы классификационной модели опорных векторов SVC? Каким образом их можно настроить?
3. Опишите параметры лучшей модели, полученной в лабораторной работе.

Лабораторная работа 6: Понижение размерности данных с помощью метода главных компонент и сингулярного разложения. Предсказание высоты дефекта в сварных швах трубопроводов АЭС с помощью линейных регрессионных моделей.

1 Объект контроля

Объектом контроля являются трубопроводы АЭС, нефтяные и газотрубопроводы диаметром - ДУ300. Основной материал - аустенитная сталь. Размер внешнего диаметра 325, толщина 16 мм. Протяженность шва 1020 мм

2 Система ПУЗК (Система полуавтоматического ультразвукового контроля)

Для проведения ультразвукового контроля (УЗК) служит установка, представленная на рисунке 1.

Рисунок 1 – Система ПУЗК

Основные функции системы ПУЗК :

- Выявление продольных и поперечных дефектов
- Определение координат и условных размеров дефекта

- Предназначена для проведения эксплуатационного контроля

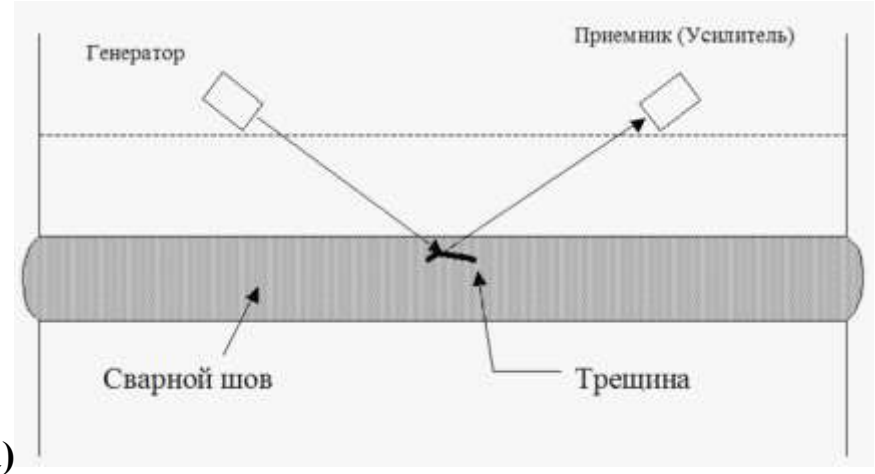
В состав системы входят 8 преобразователей, располагающихся по обе стороны сварного шва. Часть из них является генераторами, а часть приемниками (усилителями) акустического сигнала (обозначены буквами Г и У), два преобразователя совмещают эти функции.

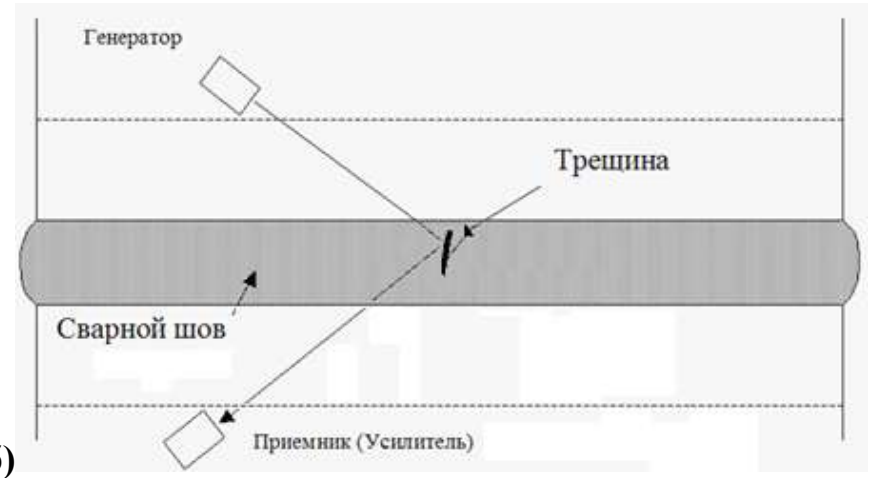
Эхо-метод

При эхо-методе преобразователи располагаются с одной стороны сварного соединения. Метод основан на том, что генератор излучает ультразвуковую волну, которая отражается от дефекта и принимается усилителем. В отсутствие дефекта сигнал на приемнике отсутствует.

Хордовая схема

а)





б)

Рисунок 2 - Схема хордового эхо-метода для:

а) продольных дефектов и б) поперечных дефектов

Раздельно-совмещенная схема

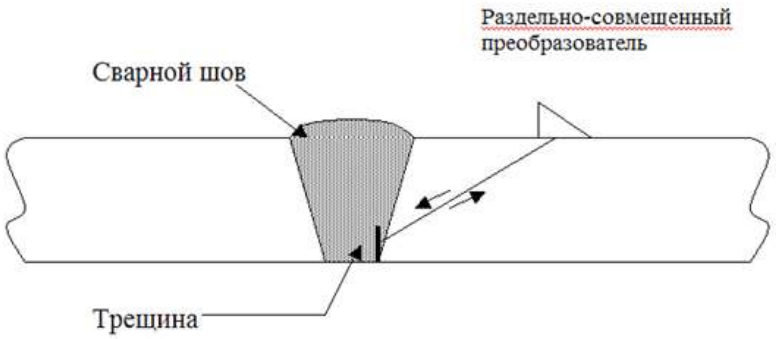


Рисунок 3 - Схема раздельно-совмещенного эхо-метода

Теневой метод

При теневом методе генератор и приемник располагаются с разных сторон шва. Если дефекта нет, волна без потерь проходит от генератора к приемнику. При наличии дефекта сигнал на приемнике ослаблен из-за рассеивания ультразвуковой волны на дефекте.

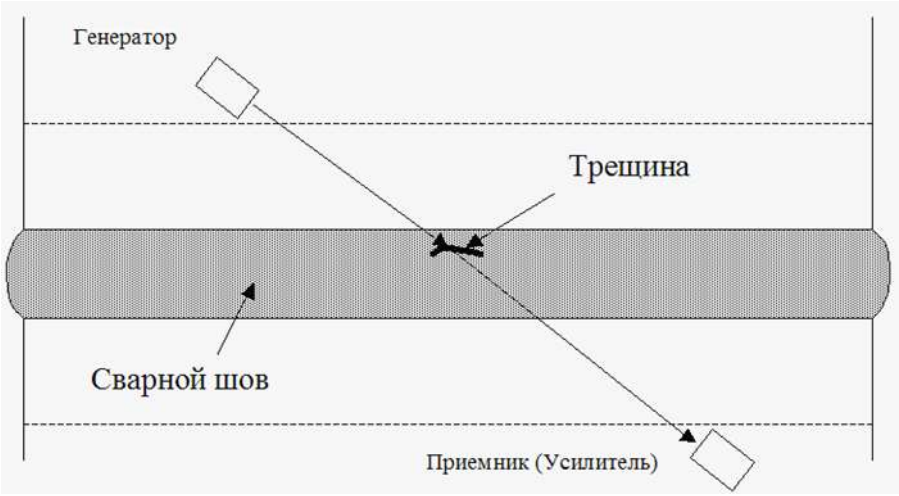


Рисунок 4 - Схема теневого метода контроля

Всего реализовано 16 различных схем прозвучивания материала сварного шва, описанные в таблице 1. Основными являются 4 схемы с использованием эхо-метода (эхо-такты, например, с генератором Г0 и приемником У0) и 4 с использованием теневого метода (теневые такты, например, Г6-У5). С их помощью осуществляется выявление продольных дефектов. Еще 2 эхо-схемы (Г2-У0 и Г0-У2) предназначены для обнаружения поперечных дефектов, которые также используют для выявления дефектов теневые схемы прозвучивания.

На рисунке 5 представлены схемы и методы прозвучивания объекта контроля, которые реализуются в блоке генераторов и приемников ультразвукового контроля, устанавливаемого на сварного соединение с помощью направляющего кольца.

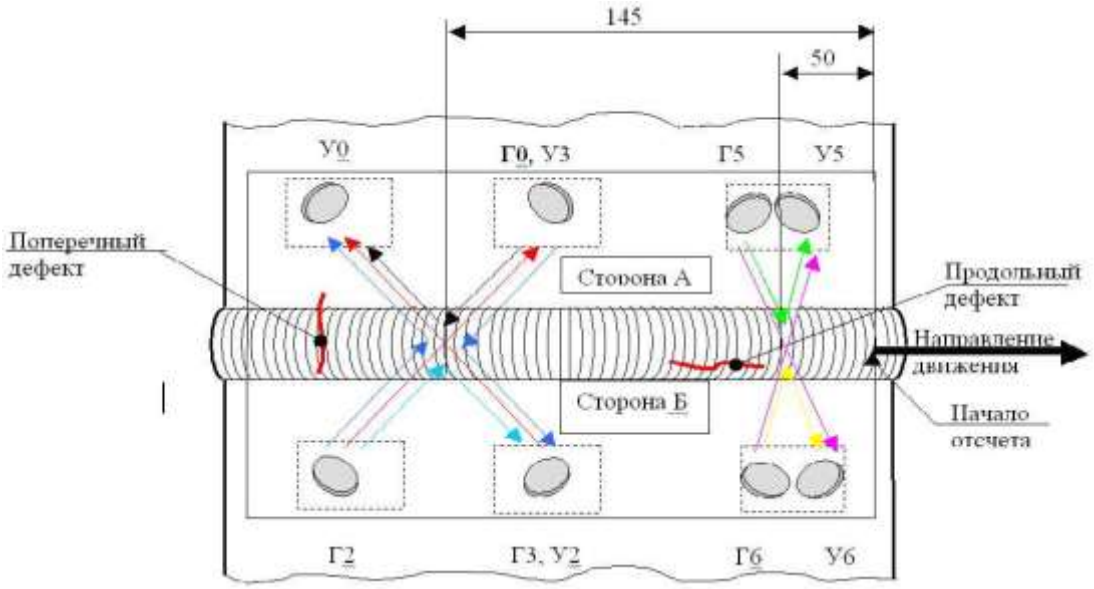


Рисунок 5 - Схема установки для проведения УЗК

Таблица 1 - Схемы прозвучивания

Такт	Генератор	Усилитель	Метод прозвучивания	Схема	Выявляемые несплошности
1	Г0	У0	Эхо-метод	Хордовая	Продольные сторона А
2	Г2	У2	Эхо-метод	Хордовая	Продольные сторона Б
3	Г5	У5	Эхо-метод	Р-С	Продольные сторона А
4	Г6	У6	Эхо-метод	Р-С	Продольные сторона Б
5	Г5	У6	Теневой метод	Р-С	Продольные сторона А
6	Г6	У5	Теневой метод	Р-С	Продольные сторона Б
7	Г0	У2	Эхо-метод	Хордовая	Поперечные
8	Г2	У0	Эхо-метод	Хордовая	Поперечные
9	Г5	У5	Эхо-Контактный м.	Р-С	Продольные сторона А
10	Г6	У6	Эхо-Контактный м.	Р-С	Продольные сторона Б
11	Г0	У0	Эхо-Контактный м.	Хордовая	Продольные сторона А
12	Г2	У2	Эхо-Контактный м.	Хордовая	Продольные сторона Б
13	Г2	У3	Теневой метод	Хордовая	Продольные сторона А

14	Г3	У0	Теневой метод	Хордовая	Продольные сторона Б
15	Г0	У2	Эхо-Контактный м.	Хордовая	Поперечные
16	Г2	У0	Эхо-Контактный м.	Хордовая	Поперечные

На случай недостаточного акустического контакта эхо-такты повторяются с усилением +6дБ (эхо-контактные) у 6 схем. Такое количество преобразователей и реализуемых с их помощью схем прозвучивания обеспечивает более надежное выявление дефектов.

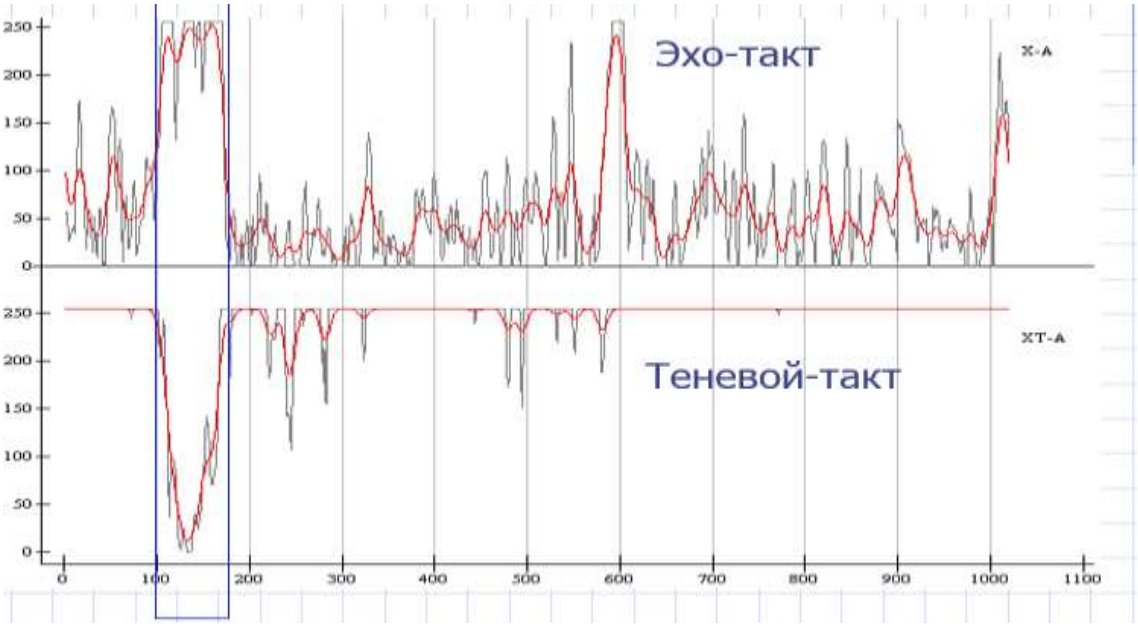
Конструктивно все преобразователи объединены в так называемый сканер, в который также входят двигатель и датчик пути. Для проведения контроля сканер с помощью специального кольца устанавливается на сварное соединение и при помощи двигателя делает один оборот вокруг трубопровода с шагом 1 мм. При этом каждый миллиметр материала шва прозвучивается по всем 16 схемам, а датчик пути измеряет пройденное расстояние. С помощью кабеля сканер соединен с ультразвуковым дефектоскопом, на который в процессе контроля передается вся полученная информация. По окончании контроля данные с дефектоскопа переносятся на персональный компьютер для дальнейшего анализа.

Алгоритм приведения данных к одной координате

Типы выявляемых дефектов(аномалий в данных)

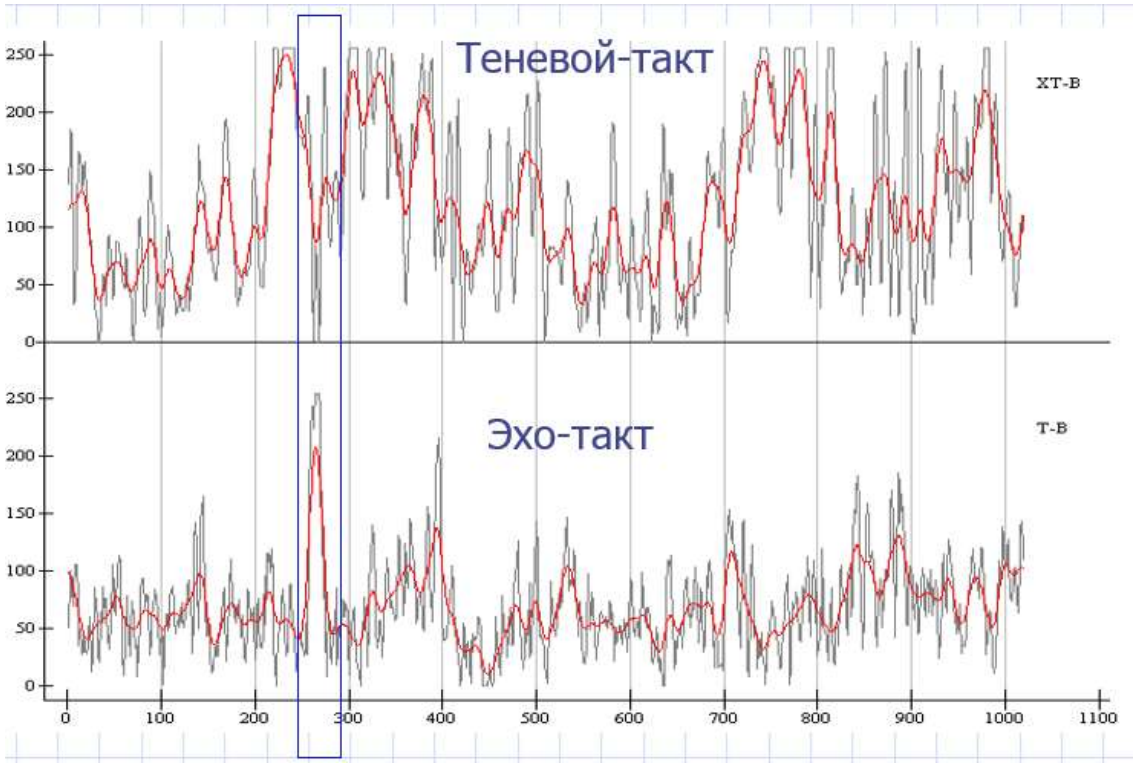
Продольные дефекты

Продольные дефекты - дефекты (трещины), расположенные вдоль оси сварного шва



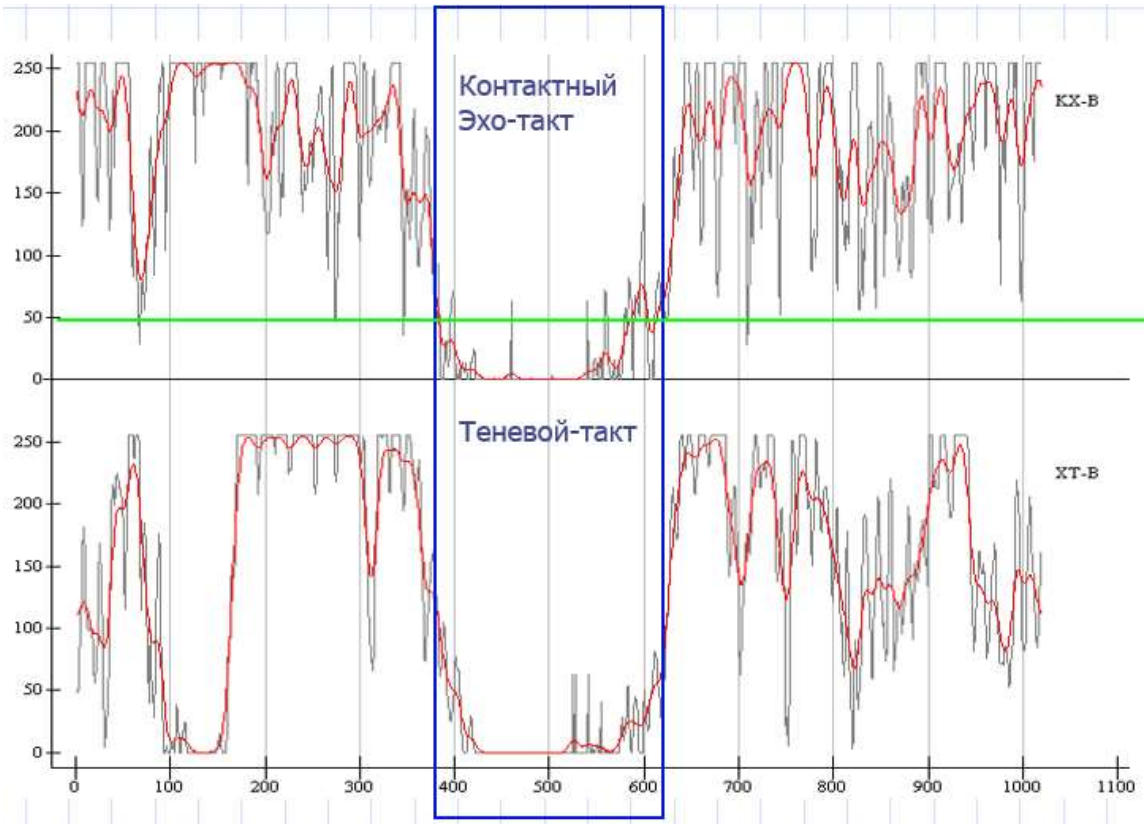
Поперечные дефекты

Поперечные дефекты - дефекты, расположенные перпендикулярно оси сварного шва.

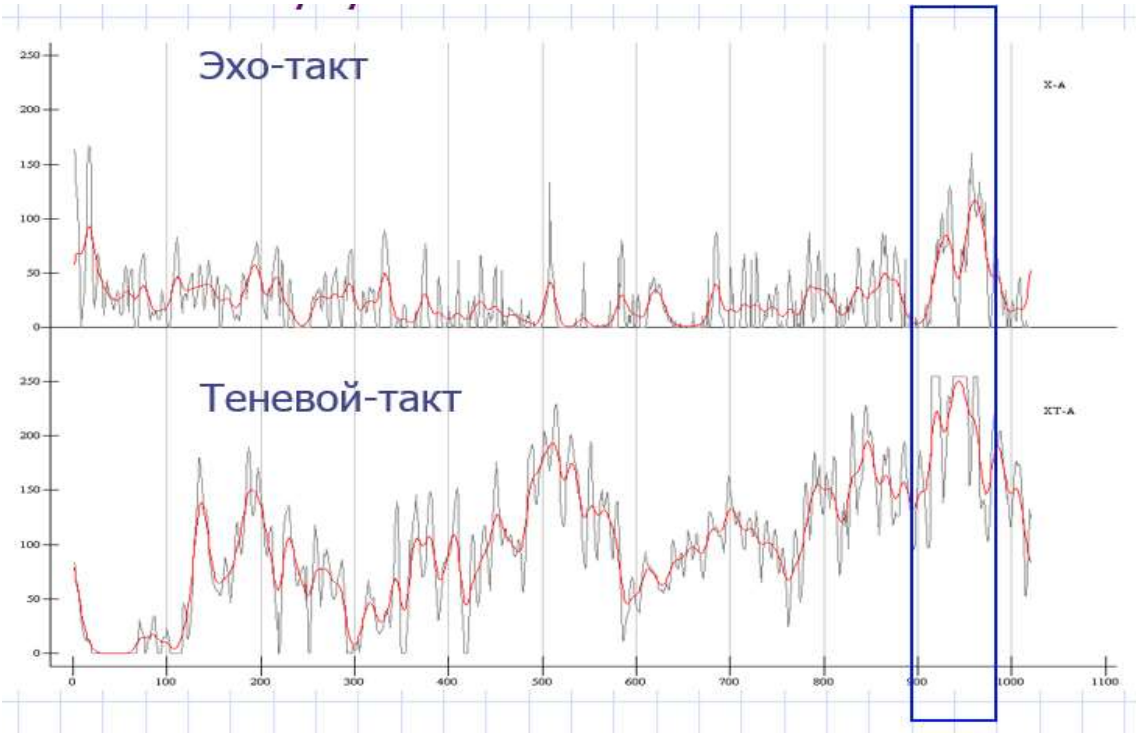


Потеря акустического контакта

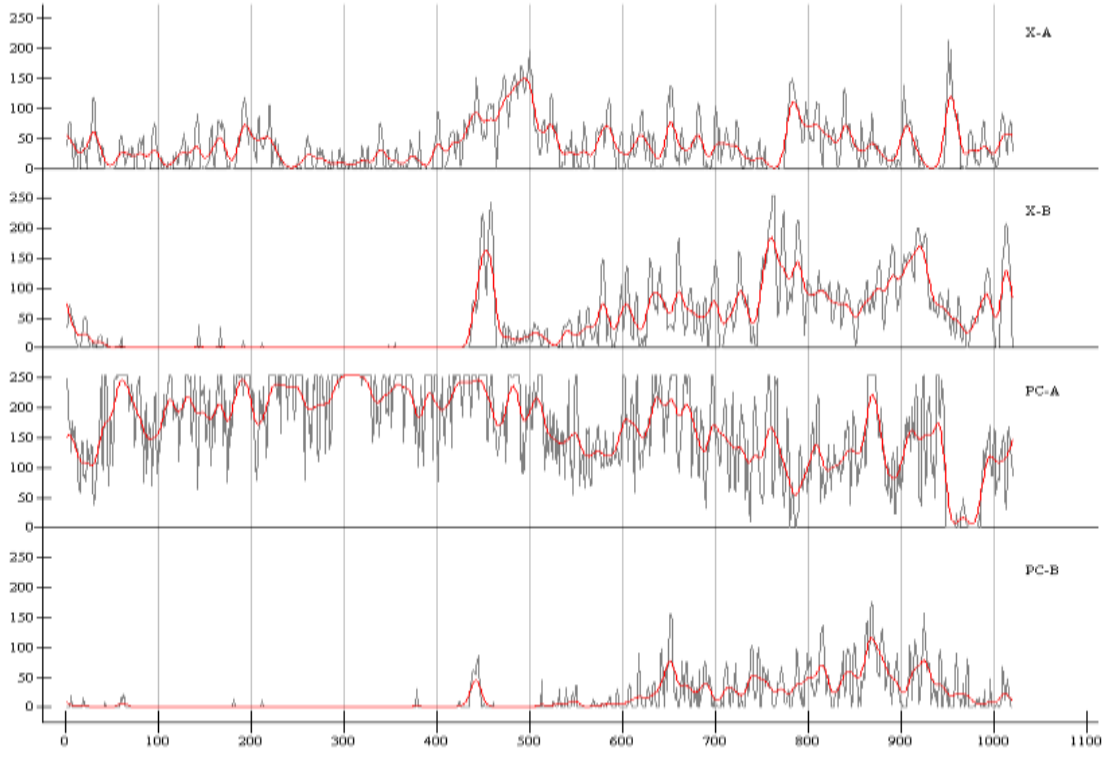
Акустический контакт – способ передачи акустического сигнала из объекта контроля в преобразователь и наоборот. Акустические волны сильно отражаются от тонких воздушных зазоров. Поэтому для передачи волн от преобразователя к объекту такие промежутки часто заполняются жидкостью.



Локальное улучшение акустического контакта



Нестационарность сигнала



3 Постановка задачи

Описанная выше система в течение нескольких лет используется на российских АЭС. Анализ результатов контроля выполняется экспертом, который выдает заключение о наличии дефектов в данном сварном

соединении и их координатах. Основным признаком дефекта является одновременное повышение уровня эхо-сигнала (пик) и падение амплитуды теневого сигнала (провал) хотя бы по одной паре тактов. Таким образом, основная задача эксперта состоит в выделении пиков и провалов сигнала на фоне помех. После определения координат дефекта, его высота определяется по величине падения теневого сигнала.

В идеале амплитуда эхо-сигнала при отсутствии дефекта должна равняться нулю, а амплитуда теневого сигнала – 255 усл.ед. При наличии дефекта должно наблюдаться обратное соотношение сигналов по эхо и теневым тактам.

В реальности, анализ сигналов затруднен наличием целого ряда мешающих факторов. Даже при отсутствии дефекта, ультразвуковая волна отражается на границах зерен структуры материала. Поэтому в сигнале всегда присутствует так называемый структурный шум. Свое влияние оказывают электрические помехи и ошибки амплитудного квантования сигналов. . Поведение сигналов УЗК существенно зависит от размера, ориентации и положения дефекта относительно измерительного блока. Наконец сильнейшее влияние на сигнал оказывает непостоянство акустического контакта датчиков и контролируемой поверхности

Таким образом, эксперт должен проводить одновременный анализ и сопоставление, в условиях шумов и мешающих факторов, 16–и сигналов, изменяющихся при изменении координат сканера. Понятны высокие требования к квалификации и опыту эксперта, которые часто недостижимы штатным персоналом лабораторий контроля металлов на АЭС. Это приводит к необходимости привлечения для контроля сотрудников организаций – разработчиков реактора и диагностического оборудования. Другими проблемами являются низкая скорость обработки результатов, субъективность оценки состояния сварного шва и влияние на нее “человеческого фактора”.

4 Результаты УЗК

Результаты УЗК сварного соединения представляют собой **файл данных**, в котором записана служебная информация (номер соединения, условия контроля и т.д.) и таблица измеренных значений сигналов. Первая строка файла является служебной и содержит информацию о номере сварного шва, о приборе контроля, дате и времени контроля, температуре и пр. В первом столбце таблицы записываются показания датчика пути (расстояние вдоль сварного шва в миллиметрах), а в остальных значения амплитуд сигналов по всем 16 схемам прозвучивания. Длина окружности трубопровода составляет 1020 мм. Для надежного контроля начального участка сканирование проводится с нахлестом от 10 до 100 мм. Амплитуда сигнала изменяется в диапазоне 0–255 условных единиц.

Пример файла с данными

Файлы с данными носят числовые названия, которые соответствуют номеру сварного шва при проведении УЗК. Форматом файлов с данными является .dat. Ниже показан пример одного такого файла.

```
0000 08-10-02 16:24:04 +33C   246   325   16   9 1 7У1      ЗОТ П2
0001 000 000 001 027 191 055 010 000 094 215 122 073 217 255 113 058
0002 000 015 008 031 134 090 009 002 116 222 080 146 179 255 105 080
0003 000 029 039 009 135 115 009 021 185 157 076 182 101 255 107 073
0004 003 029 109 000 255 164 017 033 255 121 098 186 112 255 118 095
0005 007 016 135 000 255 152 023 035 255 115 110 153 119 252 135 136
0006 016 003 087 000 255 118 025 031 255 078 127 114 119 234 139 089
0007 016 015 064 000 255 078 018 036 231 077 125 147 111 199 124 097
0008 009 028 079 000 244 065 034 045 255 096 109 186 098 224 158 110
0009 000 043 069 000 251 066 024 075 238 108 087 223 088 255 135 157
0010 001 064 041 006 255 061 000 073 178 147 094 255 067 255 079 150
0011 003 078 036 000 239 044 000 033 176 081 099 255 081 250 053 126
0012 005 088 046 000 212 037 000 054 200 027 102 255 141 255 042 149
0013 005 079 048 000 163 030 000 047 201 037 107 255 135 255 070 136
0014 015 069 032 000 160 058 001 040 162 041 129 255 113 255 089 163
0015 021 053 018 000 127 065 000 028 129 021 137 239 123 255 074 161
0016 021 045 015 000 072 028 015 042 126 000 139 225 195 242 121 146
0017 013 022 057 000 138 033 009 058 225 000 126 161 250 200 108 169
0018 008 005 096 000 159 000 034 082 255 000 108 119 255 153 155 170
0019 005 000 094 000 082 000 046 077 255 063 103 075 244 105 180 158
0020 003 000 067 000 131 007 031 042 234 005 098 036 222 066 147 150
0021 000 000 050 000 068 028 003 057 200 000 084 034 220 061 091 154
0022 000 000 069 000 041 011 000 041 242 000 063 043 214 114 068 131
0023 000 000 061 000 032 000 000 027 223 000 030 055 181 168 061 130
```

Рисунок 5 - Пример файла с данными

5 Целевая переменная

Целевую переменную нужно будет сделать из таблицы (csv формат), в столбцах которой будут указаны характеристики найденных дефектов.

Таблица дефектов

Начало дефекта	Длина дефекта	Высота дефекта	Тип дефекта	Сторона
20	30	4	L	B
130	35	5	L	A
256	29	3	L	A
310	38	7	L	B
515	25	6	L	A
830		7	T	
910		5	T	
178		3	T	

Комментарий к таблице:

Тип дефекта: L - протяженный, T - поперечный.

6 Обнаружение дефектов

Проявление дефекта в сигнале эхо-такта можно представить как увеличение уровня сигнала от некоторого начального значения, области постоянного уровня (при сканировании вдоль дефектной области) и последующим снижении уровня сигнала.

Длину дефекта определяют как разность координат конца и начала сигнала от дефекта на С-скане, то есть разность границ дефекта.

Высоту дефекта определяют по уровню падения сигнала от несплошности. Так например, если падение сигнала от 255 усл.ед. составляет 200 усл.ед. то высота дефекта находится в диапазоне от 8 мм и более. А если падение сигнала от 255 усл.ед. находится в диапазоне от 10 до 20 усл.ед. то высота дефекта будет 2-3 мм.

7 Тренировочная выборка

В качестве тренировочной последовательности будет использоваться выборка SOP, полученная в результате сканирования системой ПУЗК стандартного образца предприятия (СОП). Сканирование образца выполнялось 3 раза подряд, именно поэтому даются 3 выборки для обучения.

Выполнение лабораторной работы

1. Прочтите данные из файлов в папке train и test. Файлы Sop1, Sop2 и Sop3 являются массивами данных обучающей выборки, а target1, target2, target3 – целевыми переменными для каждого файла Sop соответственно.

Функции, которые могут пригодиться при решении: pd.read_csv()

2. Отобразите несколько первых и несколько последних записей.

Функции, которые могут пригодиться при решении: .head(), .tail().

3. Постройте гистограммы, ящики с усами и временные реализации сигналов Sop из обучающей выборки (train).
4. С помощью массивов, содержащих значения целевой переменной, создайте вектор с результатами наличия дефектов по всей длине сварного шва, состоящего из 0 (отсутствие дефекта) и 1 (наличие дефекта).

5. Примените метод понижения размерности (метод главных компонент) к исходному набору данных с частотами спектров. Визуализируйте 2 первые главные компоненты на плоскости и раскрасьте точки на графике с помощью созданного на предыдущем шаге вектора целевой переменной. (*from sklearn.decomposition import PCA)*. Не забудьте выполнить масштабирование многомерных данных перед их визуализацией на плоскости и понижением размерности с помощью функции StandardScaler из библиотеки sklearn.
6. С помощью массивов, содержащих значения целевой переменной, создайте еще один вектор, содержащий значения высоты дефектов в местах их наличия. На всех остальных интервалах поставьте нулевые значения.
7. Разбейте данные из папки train на обучающую и проверочную (валидационную) выборки в пропорции 70 на 30 с помощью функции *train_test_split()* из библиотеки *sklearn*.
8. Последовательно обучите алгоритм линейной регрессии (LinearRegression), стохастического градиентного спуска (SGDRegressor), гребневой регрессии (Ridge) и классификатор лассо (Lasso) на массиве обучающей выборки с параметрами, установленными по-умолчанию. Перечисленные алгоритмы можно загрузить используя модуль библиотеки sklearn – linear_model (например, *from sklearn.linear_model import SGDRegressor*.
9. Выполните предсказание на тестовой выборке из папки test. Оцените качество модели с помощью метрики *R-square (коэффициент детерминации) и MAE (средняя абсолютная ошибка)* из библиотеки *sklearn* модуля *metrics. (from sklearn.metrics import r2, mean_absolute_error).* Выберите наилучший алгоритм. Аргументируйте свой выбор.

10. Выполните подбор гиперпараметров наилучшей модели, выбранной на предыдущем шаге, с помощью *GridSearchCV() (from sklearn.model_selection import GridSearchCV)* с параметром кросс-валидации cv = 5. Подумайте какие параметры стоит настроить. Аргументируйте свой выбор.
11. Заново обучите наилучшую модель с подобранными гиперпараметрами на обучающей выборке и оцените качество ее работы на тестовой (метрики – MAE, R-Square).
12. Оформите отчет по лабораторной работе в формате ipynb с заголовками, комментариями, рисунками (с заголовками и названиями осей), ответами на контрольные вопросы, а также выводами о проделанной работе. Перед первым заголовком должно быть ваше ФИО и название группы. Назовите файл ФИО_lab6.ipynb и сделайте файл .pdf с таким же названием, а затем сдайте оба файла преподавателю.

Контрольные вопросы

1. Что такое система ПУЗК? Какие основные методы УЗК реализуются в данной системе?
2. Чем отличается эхо-метод от теневого и эхо-контактного метода УЗК?
3. Какие типы дефектов можно найти с помощью системы ПУЗК?

Лабораторная работа 7: Техники градиентного бустинга, бэггинга и стэкинга для решения задач классификации и регрессии. Решение задачи классификации типа дефекта в сварных швах трубопроводов АЭС.

1 Объект контроля

Объектом контроля являются трубопроводы АЭС, нефтяные и газотрубопроводы диаметром - ДУ300. Основной материал - аустенитная сталь. Размер внешнего диаметра 325, толщина 16 мм. Протяженность шва 1020 мм

2 Система ПУЗК (Система полуавтоматического ультразвукового контроля)

Для проведения ультразвукового контроля (УЗК) служит установка, представленная на рисунке 1.

Рисунок 1 – Система ПУЗК

Основные функции системы ПУЗК :

- Выявление продольных и поперечных дефектов
- Определение координат и условных размеров дефекта

- Предназначена для проведения эксплуатационного контроля

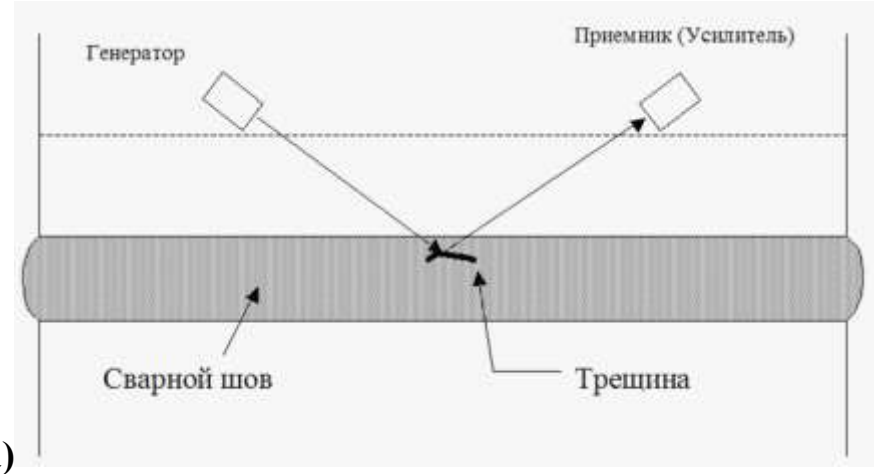
В состав системы входят 8 преобразователей, располагающихся по обе стороны сварного шва. Часть из них является генераторами, а часть приемниками (усилителями) акустического сигнала (обозначены буквами Г и У), два преобразователя совмещают эти функции.

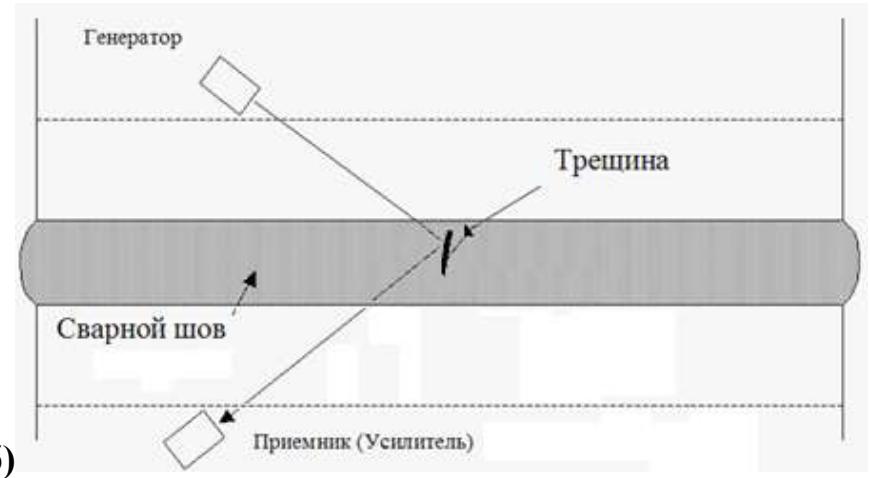
Эхо-метод

При эхо-методе преобразователи располагаются с одной стороны сварного соединения. Метод основан на том, что генератор излучает ультразвуковую волну, которая отражается от дефекта и принимается усилителем. В отсутствие дефекта сигнал на приемнике отсутствует.

Хордовая схема

а)





б)

Рисунок 2 - Схема хордового эхо-метода для:

а) продольных дефектов и б) поперечных дефектов

Раздельно-совмещенная схема

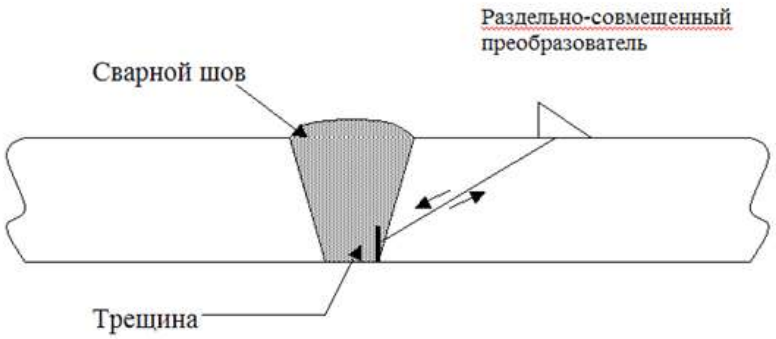


Рисунок 3 - Схема раздельно-совмещенного эхо-метода

Теневой метод

При теневом методе генератор и приемник располагаются с разных сторон шва. Если дефекта нет, волна без потерь проходит от генератора к приемнику. При наличии дефекта сигнал на приемнике ослаблен из-за рассеивания ультразвуковой волны на дефекте.

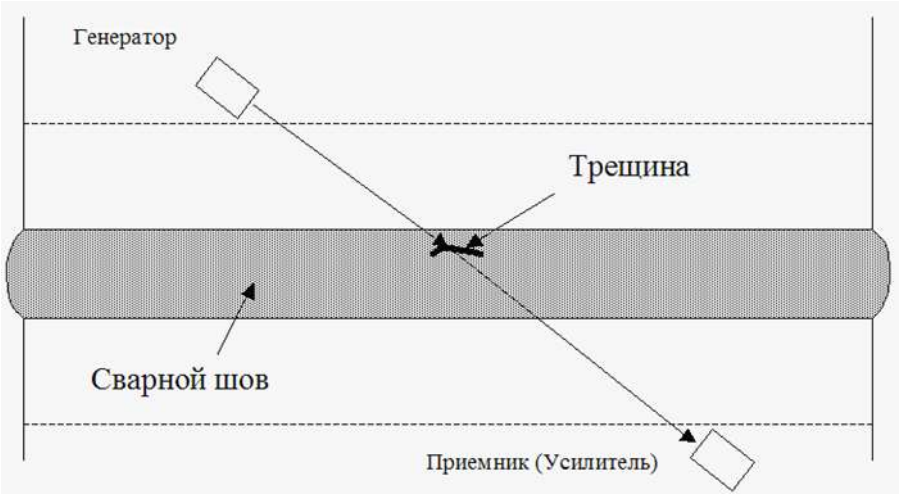


Рисунок 4 - Схема теневого метода контроля

Всего реализовано 16 различных схем прозвучивания материала сварного шва, описанные в таблице 1. Основными являются 4 схемы с использованием эхо-метода (эхо-такты, например, с генератором Г0 и приемником У0) и 4 с использованием теневого метода (теневые такты, например, Г6-У5). С их помощью осуществляется выявление продольных дефектов. Еще 2 эхо-схемы (Г2-У0 и Г0-У2) предназначены для обнаружения поперечных дефектов, которые также используют для выявления дефектов теневые схемы прозвучивания.

На рисунке 5 представлены схемы и методы прозвучивания объекта контроля, которые реализуются в блоке генераторов и приемников ультразвукового контроля, устанавливаемого на сварного соединение с помощью направляющего кольца.

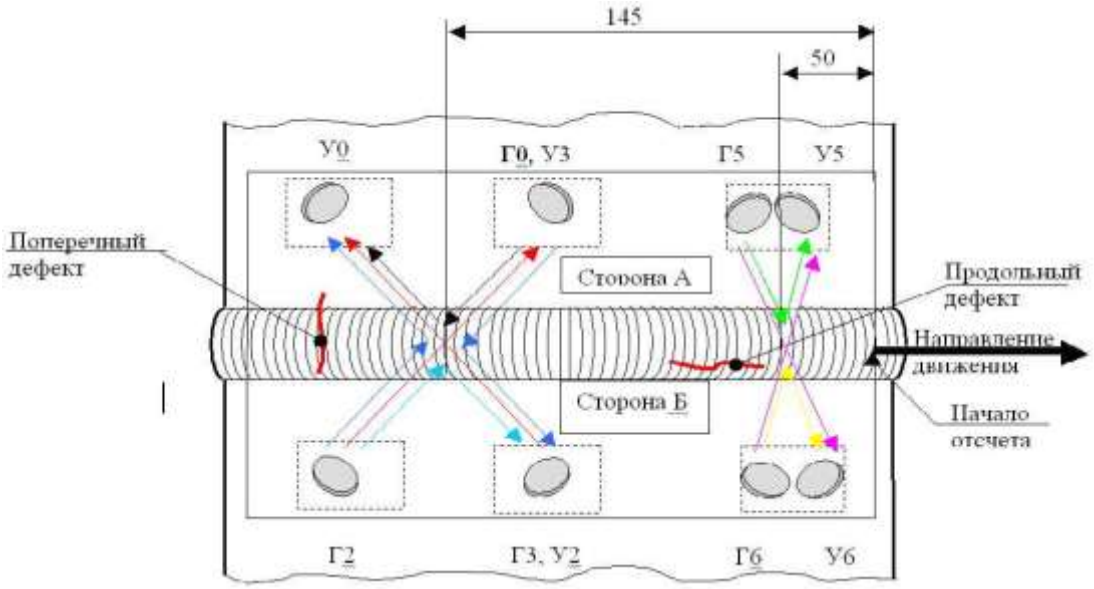


Рисунок 5 - Схема установки для проведения УЗК

Таблица 1 - Схемы прозвучивания

Такт	Генератор	Усилитель	Метод прозвучивания	Схема	Выявляемые несплошности
1	Г0	У0	Эхо-метод	Хордовая	Продольные сторона А
2	Г2	У2	Эхо-метод	Хордовая	Продольные сторона Б
3	Г5	У5	Эхо-метод	Р-С	Продольные сторона А
4	Г6	У6	Эхо-метод	Р-С	Продольные сторона Б
5	Г5	У6	Теневой метод	Р-С	Продольные сторона А
6	Г6	У5	Теневой метод	Р-С	Продольные сторона Б
7	Г0	У2	Эхо-метод	Хордовая	Поперечные
8	Г2	У0	Эхо-метод	Хордовая	Поперечные
9	Г5	У5	Эхо-Контактный м.	Р-С	Продольные сторона А
10	Г6	У6	Эхо-Контактный м.	Р-С	Продольные сторона Б
11	Г0	У0	Эхо-Контактный м.	Хордовая	Продольные сторона А
12	Г2	У2	Эхо-Контактный м.	Хордовая	Продольные сторона Б
13	Г2	У3	Теневой метод	Хордовая	Продольные сторона А

14	Г3	У0	Теневой метод	Хордовая	Продольные сторона Б
15	Г0	У2	Эхо-Контактный м.	Хордовая	Поперечные
16	Г2	У0	Эхо-Контактный м.	Хордовая	Поперечные

На случай недостаточного акустического контакта эхо-такты повторяются с усилением +6дБ (эхо-контактные) у 6 схем. Такое количество преобразователей и реализуемых с их помощью схем прозвучивания обеспечивает более надежное выявление дефектов.

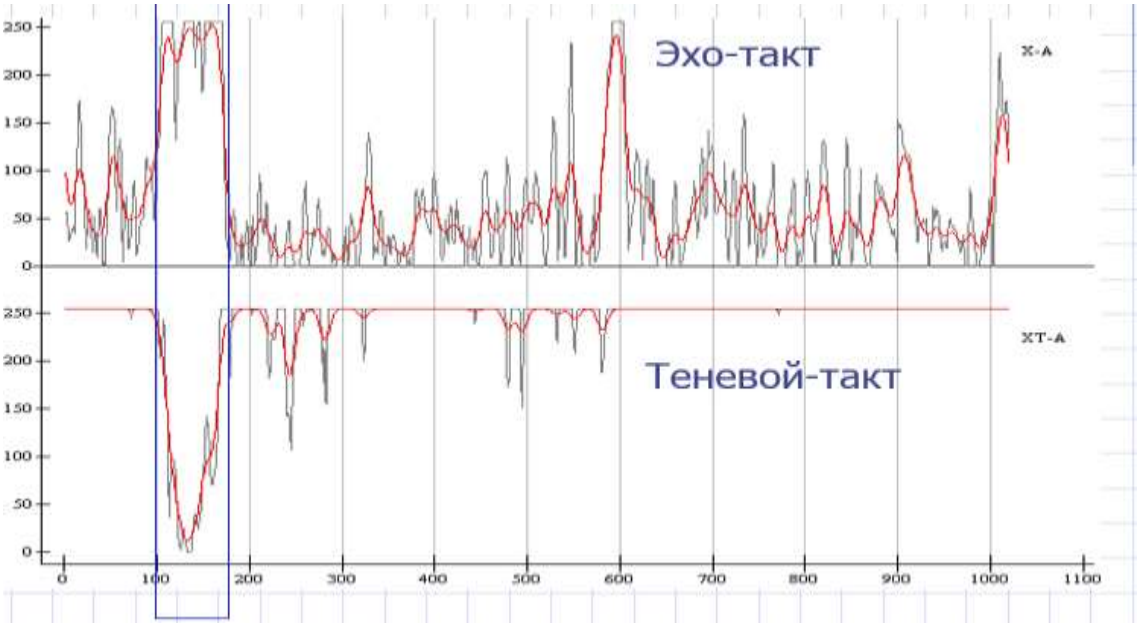
Конструктивно все преобразователи объединены в так называемый сканер, в который также входят двигатель и датчик пути. Для проведения контроля сканер с помощью специального кольца устанавливается на сварное соединение и при помощи двигателя делает один оборот вокруг трубопровода с шагом 1 мм. При этом каждый миллиметр материала шва прозвучивается по всем 16 схемам, а датчик пути измеряет пройденное расстояние. С помощью кабеля сканер соединен с ультразвуковым дефектоскопом, на который в процессе контроля передается вся полученная информация. По окончании контроля данные с дефектоскопа переносятся на персональный компьютер для дальнейшего анализа.

Алгоритм приведения данных к одной координате

Типы выявляемых дефектов(аномалий в данных)

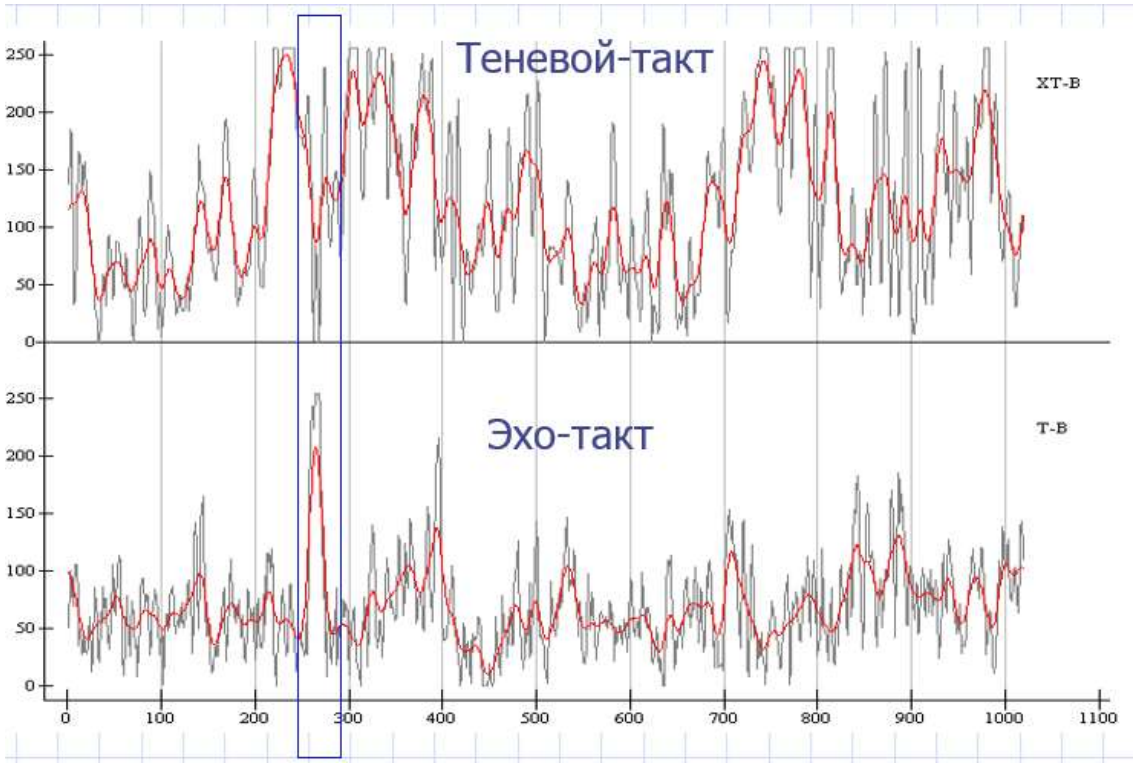
Продольные дефекты

Продольные дефекты - дефекты (трещины), расположенные вдоль оси сварного шва



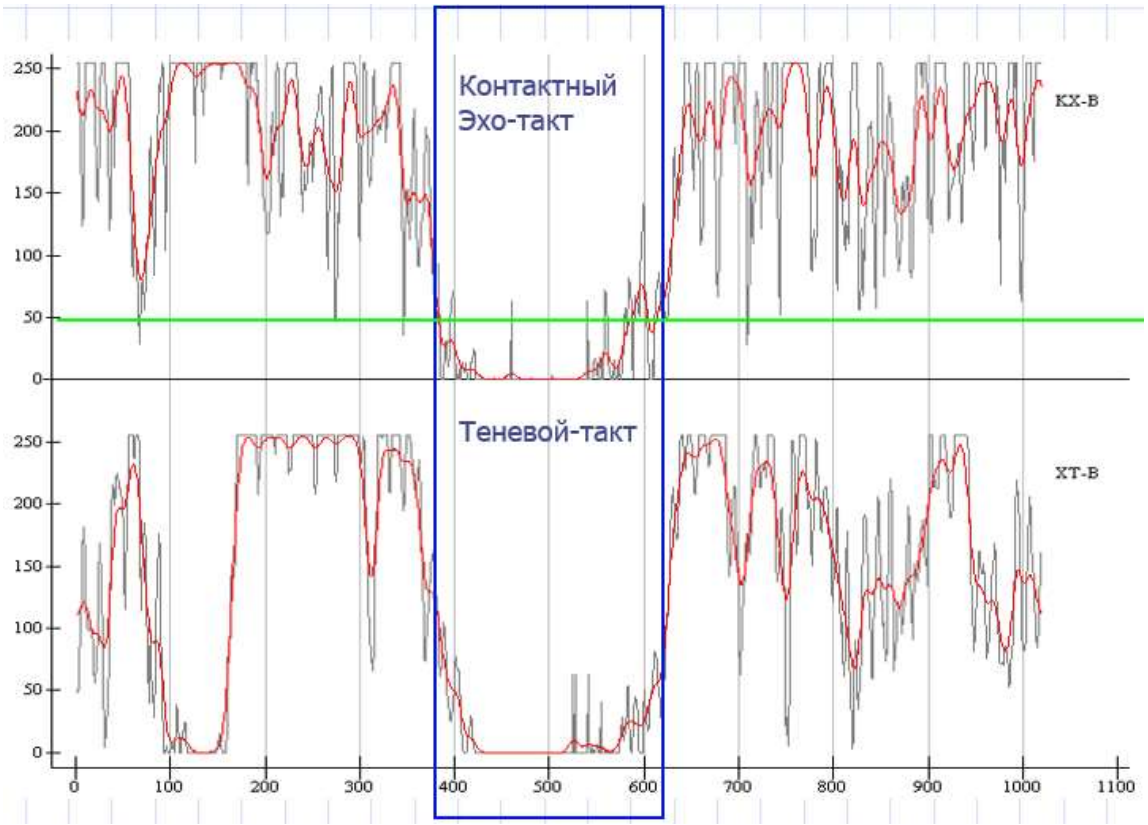
Поперечные дефекты

Поперечные дефекты - дефекты, расположенные перпендикулярно оси сварного шва.

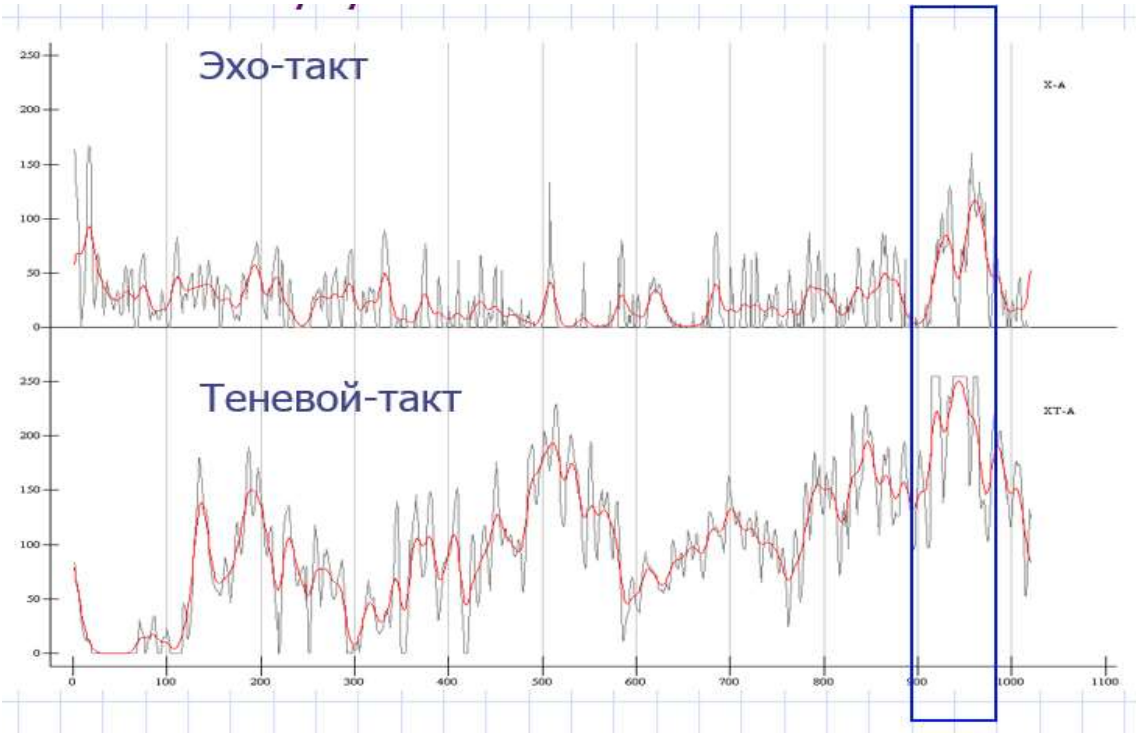


Потеря акустического контакта

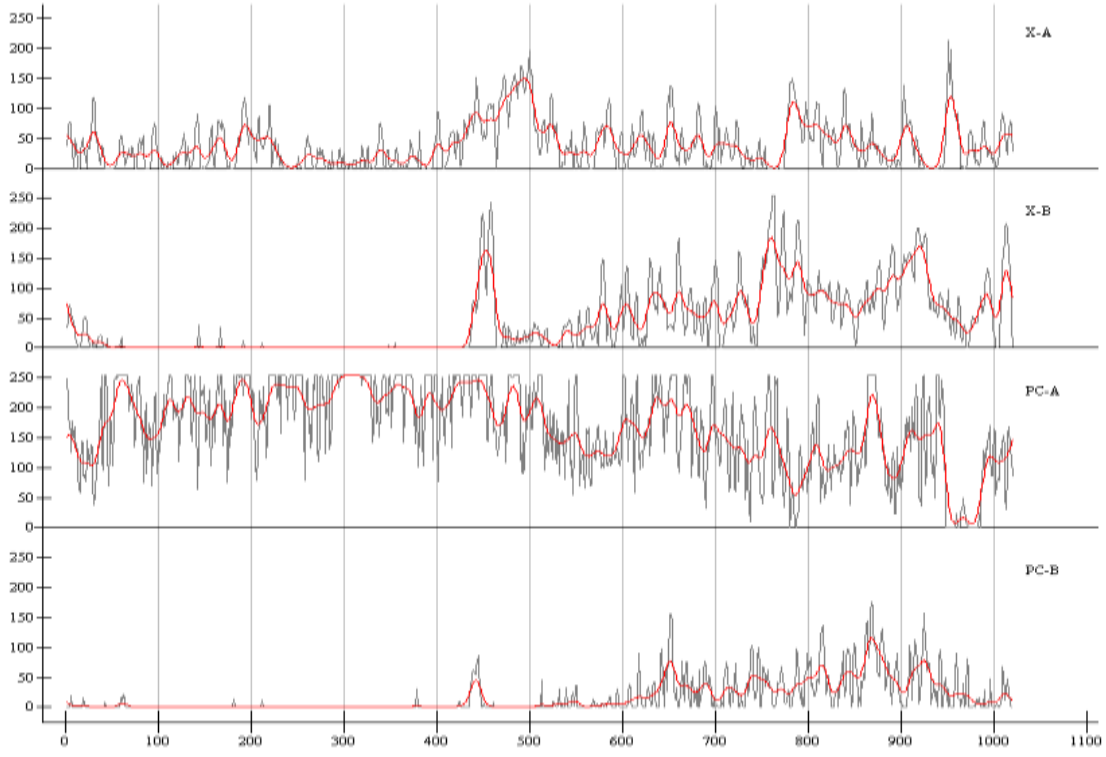
Акустический контакт – способ передачи акустического сигнала из объекта контроля в преобразователь и наоборот. Акустические волны сильно отражаются от тонких воздушных зазоров. Поэтому для передачи волн от преобразователя к объекту такие промежутки часто заполняются жидкостью.



Локальное улучшение акустического контакта



Нестационарность сигнала



3 Постановка задачи

Описанная выше система в течение нескольких лет используется на российских АЭС. Анализ результатов контроля выполняется экспертом, который выдает заключение о наличии дефектов в данном сварном

соединении и их координатах. Основным признаком дефекта является одновременное повышение уровня эхо-сигнала (пик) и падение амплитуды теневого сигнала (провал) хотя бы по одной паре тактов. Таким образом, основная задача эксперта состоит в выделении пиков и провалов сигнала на фоне помех. После определения координат дефекта, его высота определяется по величине падения теневого сигнала.

В идеале амплитуда эхо-сигнала при отсутствии дефекта должна равняться нулю, а амплитуда теневого сигнала – 255 усл.ед. При наличии дефекта должно наблюдаться обратное соотношение сигналов по эхо и теневым тактам.

В реальности, анализ сигналов затруднен наличием целого ряда мешающих факторов. Даже при отсутствии дефекта, ультразвуковая волна отражается на границах зерен структуры материала. Поэтому в сигнале всегда присутствует так называемый структурный шум. Свое влияние оказывают электрические помехи и ошибки амплитудного квантования сигналов. . Поведение сигналов УЗК существенно зависит от размера, ориентации и положения дефекта относительно измерительного блока. Наконец сильнейшее влияние на сигнал оказывает непостоянство акустического контакта датчиков и контролируемой поверхности

Таким образом, эксперт должен проводить одновременный анализ и сопоставление, в условиях шумов и мешающих факторов, 16–и сигналов, изменяющихся при изменении координат сканера. Понятны высокие требования к квалификации и опыту эксперта, которые часто недостижимы штатным персоналом лабораторий контроля металлов на АЭС. Это приводит к необходимости привлечения для контроля сотрудников организаций – разработчиков реактора и диагностического оборудования. Другими проблемами являются низкая скорость обработки результатов, субъективность оценки состояния сварного шва и влияние на нее “человеческого фактора”.

4 Результаты УЗК

Результаты УЗК сварного соединения представляют собой **файл данных**, в котором записана служебная информация (номер соединения, условия контроля и т.д.) и таблица измеренных значений сигналов. Первая строка файла является служебной и содержит информацию о номере сварного шва, о приборе контроля, дате и времени контроля, температуре и пр. В первом столбце таблицы записываются показания датчика пути (расстояние вдоль сварного шва в миллиметрах), а в остальных значения амплитуд сигналов по всем 16 схемам прозвучивания. Длина окружности трубопровода составляет 1020 мм. Для надежного контроля начального участка сканирование проводится с нахлестом от 10 до 100 мм. Амплитуда сигнала изменяется в диапазоне 0–255 условных единиц.

Пример файла с данными

Файлы с данными носят числовые названия, которые соответствуют номеру сварного шва при проведении УЗК. Форматом файлов с данными является .dat. Ниже показан пример одного такого файла.

```
0000 08-10-02 16:24:04 +33C   246   325   16   9 1 7У1      ЗОТ П2
0001 000 000 001 027 191 055 010 000 094 215 122 073 217 255 113 058
0002 000 015 008 031 134 090 009 002 116 222 080 146 179 255 105 080
0003 000 029 039 009 135 115 009 021 185 157 076 182 101 255 107 073
0004 003 029 109 000 255 164 017 033 255 121 098 186 112 255 118 095
0005 007 016 135 000 255 152 023 035 255 115 110 153 119 252 135 136
0006 016 003 087 000 255 118 025 031 255 078 127 114 119 234 139 089
0007 016 015 064 000 255 078 018 036 231 077 125 147 111 199 124 097
0008 009 028 079 000 244 065 034 045 255 096 109 186 098 224 158 110
0009 000 043 069 000 251 066 024 075 238 108 087 223 088 255 135 157
0010 001 064 041 006 255 061 000 073 178 147 094 255 067 255 079 150
0011 003 078 036 000 239 044 000 033 176 081 099 255 081 250 053 126
0012 005 088 046 000 212 037 000 054 200 027 102 255 141 255 042 149
0013 005 079 048 000 163 030 000 047 201 037 107 255 135 255 070 136
0014 015 069 032 000 160 058 001 040 162 041 129 255 113 255 089 163
0015 021 053 018 000 127 065 000 028 129 021 137 239 123 255 074 161
0016 021 045 015 000 072 028 015 042 126 000 139 225 195 242 121 146
0017 013 022 057 000 138 033 009 058 225 000 126 161 250 200 108 169
0018 008 005 096 000 159 000 034 082 255 000 108 119 255 153 155 170
0019 005 000 094 000 082 000 046 077 255 063 103 075 244 105 180 158
0020 003 000 067 000 131 007 031 042 234 005 098 036 222 066 147 150
0021 000 000 050 000 068 028 003 057 200 000 084 034 220 061 091 154
0022 000 000 069 000 041 011 000 041 242 000 063 043 214 114 068 131
0023 000 000 061 000 032 000 000 027 223 000 030 055 181 168 061 130
```

Рисунок 5 - Пример файла с данными

5 Целевая переменная

Целевую переменную нужно будет сделать из таблицы (csv формат), в столбцах которой будут указаны характеристики найденных дефектов.

Таблица дефектов

Начало дефекта	Длина дефекта	Высота дефекта	Тип дефекта	Сторона
20	30	4	L	B
130	35	5	L	A
256	29	3	L	A
310	38	7	L	B
515	25	6	L	A
830		7	T	
910		5	T	
178		3	T	

Комментарий к таблице:

Тип дефекта: L - протяженный, T - поперечный.

6 Обнаружение дефектов

Проявление дефекта в сигнале эхо-такта можно представить как увеличение уровня сигнала от некоторого начального значения, области постоянного уровня (при сканировании вдоль дефектной области) и последующим снижении уровня сигнала.

Длину дефекта определяют как разность координат конца и начала сигнала от дефекта на С-скане, то есть разность границ дефекта.

Высоту дефекта определяют по уровню падения сигнала от несплошности. Так например, если падение сигнала от 255 усл.ед. составляет 200 усл.ед. то высота дефекта находится в диапазоне от 8 мм и более. А если падение сигнала от 255 усл.ед. находится в диапазоне от 10 до 20 усл.ед. то высота дефекта будет 2-3 мм.

7 Тренировочная выборка

В качестве тренировочной последовательности будет использоваться выборка SOP, полученная в результате сканирования системой ПУЗК стандартного образца предприятия (СОП). Сканирование образца выполнялось 3 раза подряд, именно поэтому даются 3 выборки для обучения.

Выполнение лабораторной работы

1. Прочтите данные из файлов в папке train и test. Файлы Sop1, Sop2 и Sop3 являются массивами данных обучающей выборки, а target1, target2, target3 – целевыми переменными для каждого файла Sop соответственно.

Функции, которые могут пригодиться при решении: pd.read_csv()

2. Отобразите несколько первых и несколько последних записей.

Функции, которые могут пригодиться при решении: .head(), .tail().

3. С помощью массивов, содержащих значения целевой переменной, создайте еще один вектор, содержащий названия типов дефектов в местах их наличия. На всех остальных интервалах поставьте нулевые значения.
4. Разбейте данные из папки train на обучающую и проверочную (валидационную) выборки в пропорции 70 на 30 с помощью функции *train_test_split()* из библиотеки *sklearn*.

5. Последовательно обучите алгоритм случайного леса (RandomForestClassifier), градиентного бустинга (GradientBoostingClassifier), стекинг-классификатора (StackingClassifier) на массиве обучающей выборки с параметрами, установленными по-умолчанию. Перечисленные алгоритмы можно загрузить используя модуль библиотеки sklearn – ensemble (например, *from sklearn.ensemble import StackingClassifier)*.
6. Выполните предсказание на тестовой выборке из папки test. Оцените качество работы моделей с помощью метрики *accuracy* и *classification report* из библиотеки *sklearn* модуля *metrics*. Выберите наилучший алгоритм и аргументируйте свой выбор.
7. Выполните подбор гиперпараметров наилучшей модели, выбранной на предыдущем шаге, с помощью *GridSearchCV() (from sklearn.model_selection import GridSearchCV)* с параметром кросс-валидации cv = 5. Подумайте какие параметры стоит настроить. Аргументируйте свой выбор.
8. Заново обучите наилучшую модель с подобранными гиперпараметрами на обучающей выборке и оцените качество ее работы на тестовой.
9. Оформите отчет по лабораторной работе в формате ipynb с заголовками, комментариями, рисунками (с заголовками и названиями осей), ответами на контрольные вопросы, а также выводами о проделанной работе. Перед первым заголовком должно быть ваше ФИО и название группы. Назовите файл ФИО_lab7.ipynb и сделайте файл .pdf с таким же названием, а затем сдайте оба файла преподавателю.

Контрольные вопросы

1. Что такое алгоритм «Случайного леса»? Чем он отличается от обычного дерева решений?
2. Что такое алгоритм градиентного бустинга?
3. Что означает техника стекинга в машинном обучении?

Лабораторная работа 8: Наивный байесовский классификатор для определения стороны дефекта в сварных швах трубопроводов АЭС.

1 Объект контроля

Объектом контроля являются трубопроводы АЭС, нефтяные и газотрубопроводы диаметром - ДУ300. Основной материал - аустенитная сталь. Размер внешнего диаметра 325, толщина 16 мм. Протяженность шва 1020 мм

2 Система ПУЗК (Система полуавтоматического ультразвукового контроля)

Для проведения ультразвукового контроля (УЗК) служит установка, представленная на рисунке 1.

Рисунок 1 – Система ПУЗК

Основные функции системы ПУЗК :

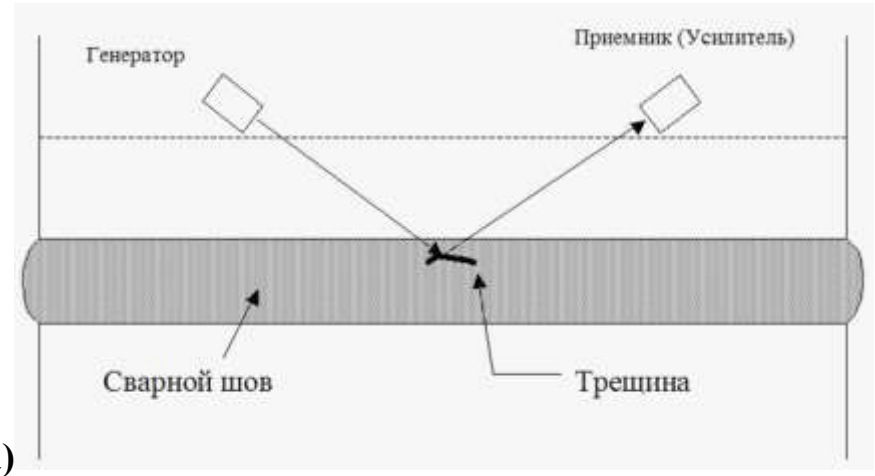
- Выявление продольных и поперечных дефектов
- Определение координат и условных размеров дефекта
- Предназначена для проведения эксплуатационного контроля

В состав системы входят 8 преобразователей, располагающихся по обе стороны сварного шва. Часть из них является генераторами, а часть приемниками (усилителями) акустического сигнала (обозначены буквами Г и У), два преобразователя совмещают эти функции.

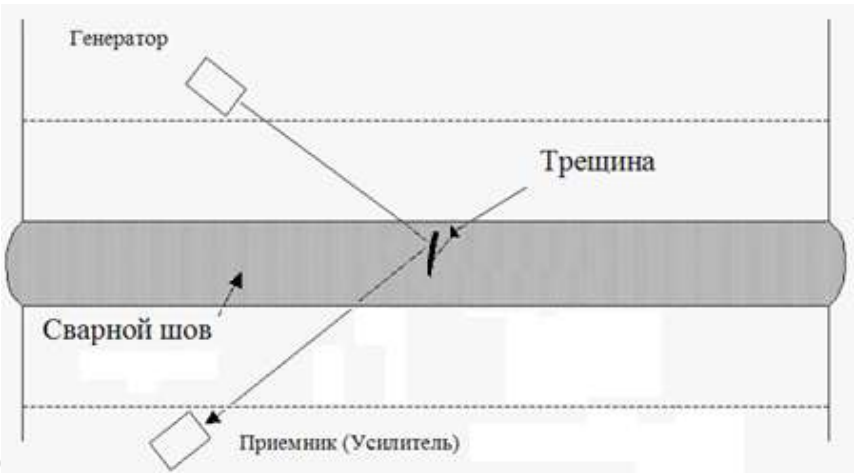
Эхо-метод

При эхо-методе преобразователи располагаются с одной стороны сварного соединения. Метод основан на том, что генератор излучает ультразвуковую волну, которая отражается от дефекта и принимается усилителем. В отсутствие дефекта сигнал на приемнике отсутствует.

Хордовая схема



а)



б)

Рисунок 2 - Схема хордового эхо-метода для:

а) продольных дефектов и б) поперечных дефектов

Раздельно-совмещенная схема

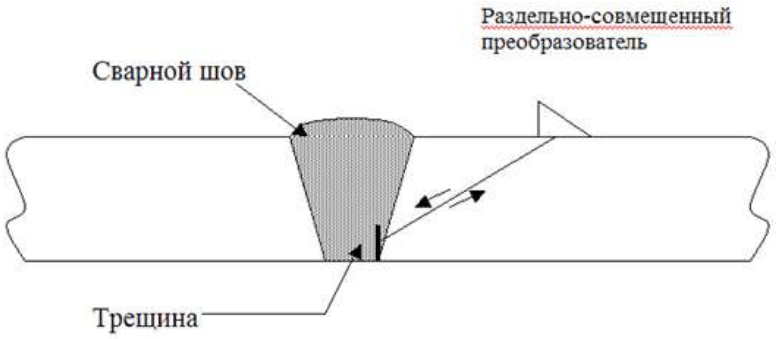


Рисунок 3 - Схема раздельно-совмещенного эхо-метода

Теневой метод

При теневом методе генератор и приемник располагаются с разных сторон шва. Если дефекта нет, волна без потерь проходит от генератора к приемнику. При наличии дефекта сигнал на приемнике ослаблен из-за рассеивания ультразвуковой волны на дефекте.

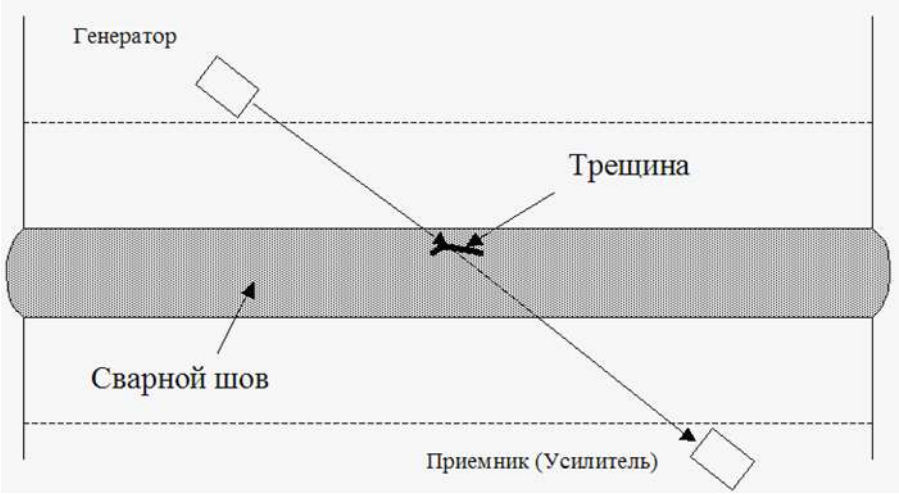


Рисунок 4 - Схема теневого метода контроля

Всего реализовано 16 различных схем прозвучивания материала сварного шва, описанные в таблице 1. Основными являются 4 схемы с использованием эхо-метода (эхо-такты, например, с генератором Г0 и приемником У0) и 4 с использованием теневого метода (теневые такты, например, Г6-У5). С их помощью осуществляется выявление продольных дефектов. Еще 2 эхо-схемы (Г2-У0 и Г0-У2) предназначены для обнаружения поперечных дефектов, которые также используют для выявления дефектов теневые схемы прозвучивания.

На рисунке 5 представлены схемы и методы прозвучивания объекта контроля, которые реализуются в блоке генераторов и приемников ультразвукового контроля, устанавливаемого на сварного соединение с помощью направляющего кольца.

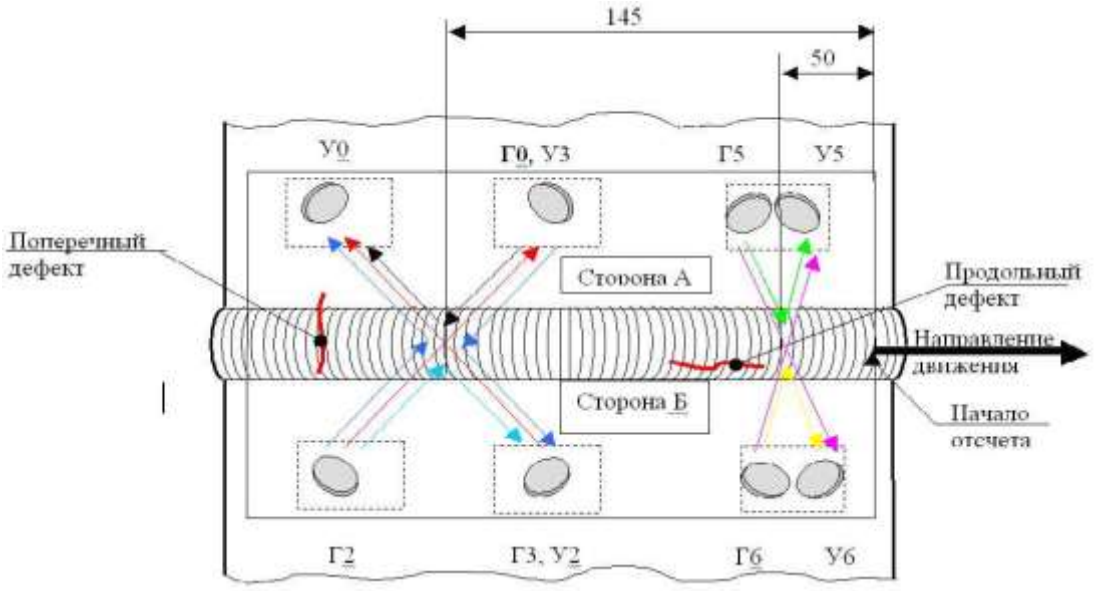


Рисунок 5 - Схема установки для проведения УЗК

Таблица 1 - Схемы прозвучивания

Такт	Генератор	Усилитель	Метод прозвучивания	Схема	Выявляемые несплошности

1	Г0	У0	Эхо-метод	Хордовая	Продольные сторона А
2	Г2	У2	Эхо-метод	Хордовая	Продольные сторона Б
3	Г5	У5	Эхо-метод	Р-С	Продольные сторона А
4	Г6	У6	Эхо-метод	Р-С	Продольные сторона Б
5	Г5	У6	Теневой метод	Р-С	Продольные сторона А
6	Г6	У5	Теневой метод	Р-С	Продольные сторона Б
7	Г0	У2	Эхо-метод	Хордовая	Поперечные
8	Г2	У0	Эхо-метод	Хордовая	Поперечные
9	Г5	У5	Эхо-Контактный м.	Р-С	Продольные сторона А
10	Г6	У6	Эхо-Контактный м.	Р-С	Продольные сторона Б
11	Г0	У0	Эхо-Контактный м.	Хордовая	Продольные сторона А
12	Г2	У2	Эхо-Контактный м.	Хордовая	Продольные сторона Б
13	Г2	У3	Теневой метод	Хордовая	Продольные сторона А
14	Г3	У0	Теневой метод	Хордовая	Продольные сторона Б
15	Г0	У2	Эхо-Контактный м.	Хордовая	Поперечные
16	Г2	У0	Эхо-Контактный м.	Хордовая	Поперечные

На случай недостаточного акустического контакта эхо-такты повторяются с усилением +6дБ (эхо-контактные) у 6 схем. Такое количество преобразователей и реализуемых с их помощью схем прозвучивания обеспечивает более надежное выявление дефектов.

Конструктивно все преобразователи объединены в так называемый сканер, в который также входят двигатель и датчик пути. Для проведения контроля сканер с помощью специального кольца устанавливается на сварное соединение и при помощи двигателя делает один оборот вокруг трубопровода с шагом 1 мм. При этом каждый миллиметр материала шва прозвучивается по всем 16 схемам, а датчик пути измеряет пройденное расстояние. С помощью кабеля сканер соединен с ультразвуковым дефектоскопом, на который в процессе контроля передается вся полученная информация. По окончании

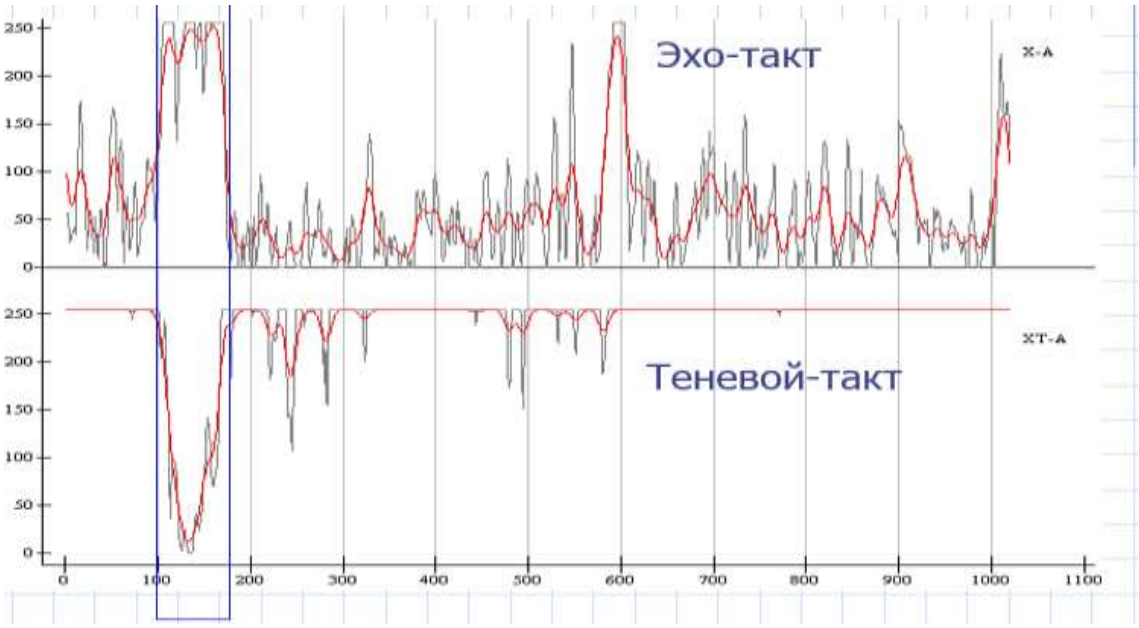
контроля данные с дефектоскопа переносятся на персональный компьютер для дальнейшего анализа.

Алгоритм приведения данных к одной координате

Типы выявляемых дефектов(аномалий в данных)

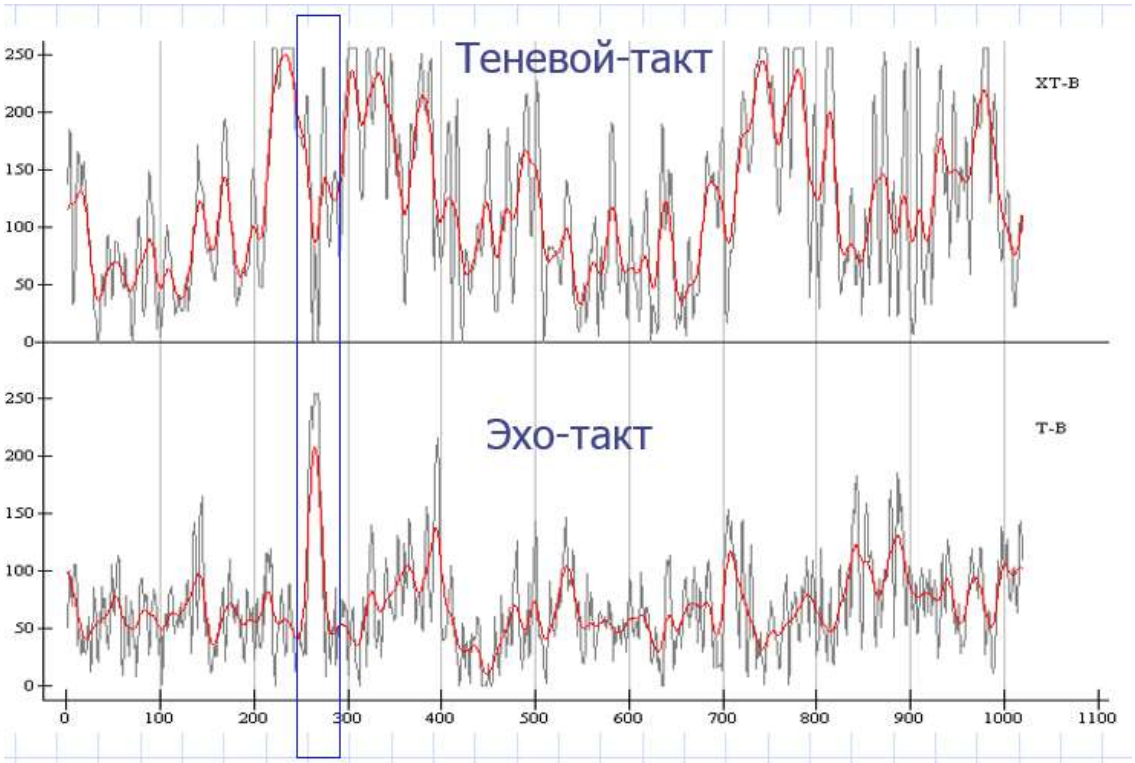
Продольные дефекты

Продольные дефекты - дефекты (трещины), расположенные вдоль оси сварного шва



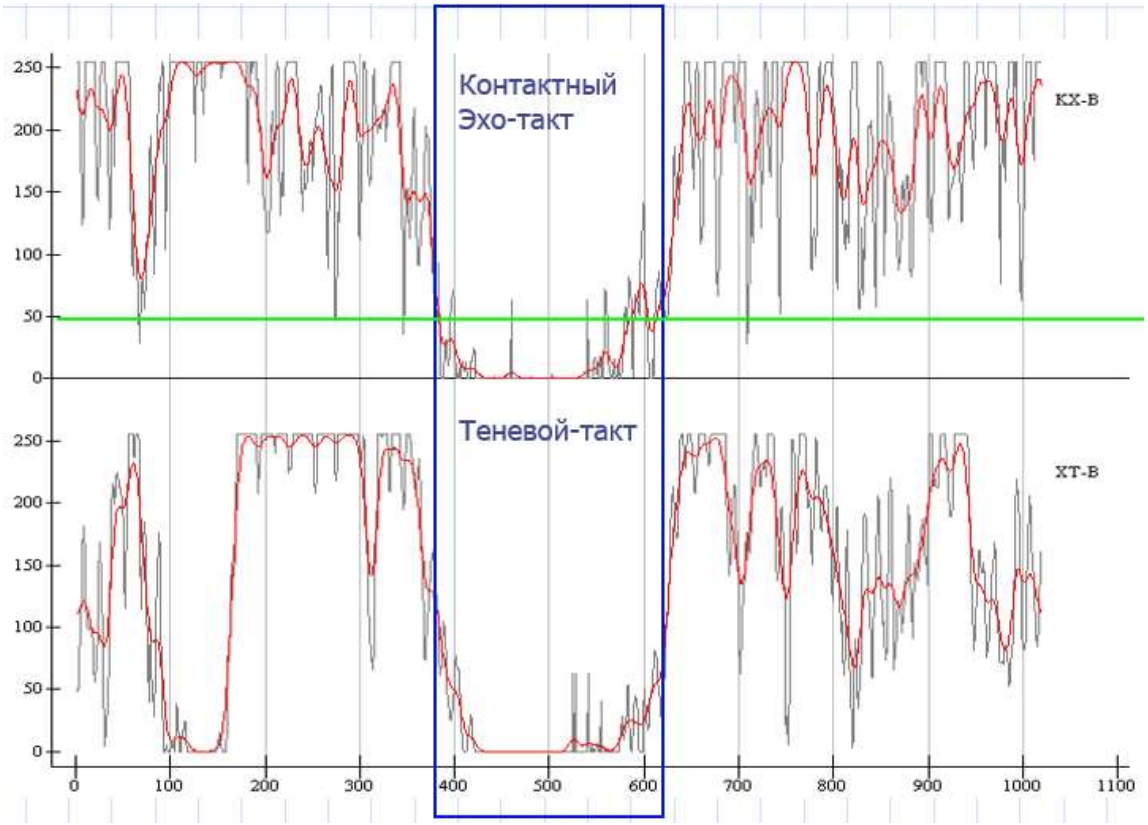
Поперечные дефекты

Поперечные дефекты - дефекты, расположенные перпендикулярно оси сварного шва.

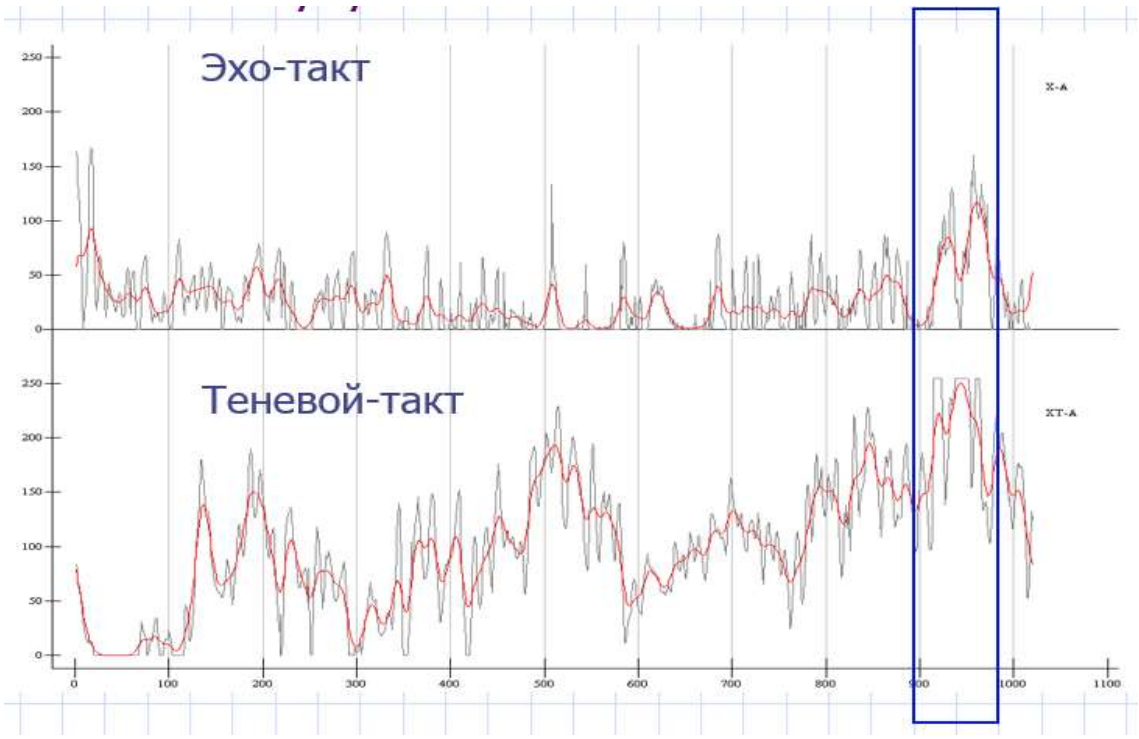


Потеря акустического контакта

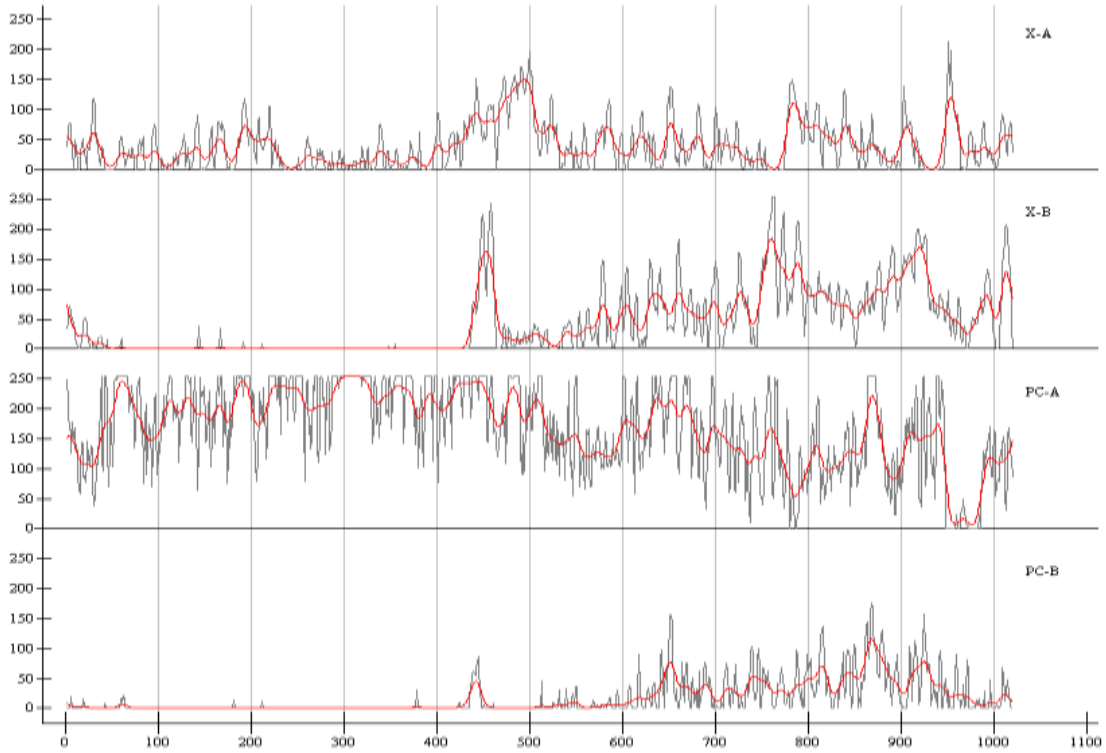
Акустический контакт – способ передачи акустического сигнала из объекта контроля в преобразователь и наоборот. Акустические волны сильно отражаются от тонких воздушных зазоров. Поэтому для передачи волн от преобразователя к объекту такие промежутки часто заполняются жидкостью.



Локальное улучшение акустического контакта



Нестационарность сигнала



3 Постановка задачи

Описанная выше система в течение нескольких лет используется на российских АЭС. Анализ результатов контроля выполняется экспертом, который выдает заключение о наличии дефектов в данном сварном соединении и их координатах. Основным признаком дефекта является одновременное повышение уровня эхо-сигнала (пик) и падение амплитуды теневого сигнала (провал) хотя бы по одной паре тактов. Таким образом, основная задача эксперта состоит в выделении пиков и провалов сигнала на фоне помех. После определения координат дефекта, его высота определяется по величине падения теневого сигнала.

В идеале амплитуда эхо-сигнала при отсутствии дефекта должна равняться нулю, а амплитуда теневого сигнала – 255 усл.ед. При наличии дефекта должно наблюдаться обратное соотношение сигналов по эхо и теневым тактам.

В реальности, анализ сигналов затруднен наличием целого ряда мешающих факторов. Даже при отсутствии дефекта, ультразвуковая волна отражается на границах зерен структуры материала. Поэтому в сигнале всегда

присутствует так называемый структурный шум. Свое влияние оказывают электрические помехи и ошибки амплитудного квантования сигналов. . Поведение сигналов УЗК существенно зависит от размера, ориентации и положения дефекта относительно измерительного блока. Наконец сильнейшее влияние на сигнал оказывает непостоянство акустического контакта датчиков и контролируемой поверхности

Таким образом, эксперт должен проводить одновременный анализ и сопоставление, в условиях шумов и мешающих факторов, 16–и сигналов, изменяющихся при изменении координат сканера. Понятны высокие требования к квалификации и опыту эксперта, которые часто недостижимы штатным персоналом лабораторий контроля металлов на АЭС. Это приводит к необходимости привлечения для контроля сотрудников организаций – разработчиков реактора и диагностического оборудования. Другими проблемами являются низкая скорость обработки результатов, субъективность оценки состояния сварного шва и влияние на нее “человеческого фактора”.

4 Результаты УЗК

Результаты УЗК сварного соединения представляют собой **файл данных**, в котором записана служебная информация (номер соединения, условия контроля и т.д.) и таблица измеренных значений сигналов. Первая строка файла является служебной и содержит информацию о номере сварного шва, о приборе контроля, дате и времени контроля, температуре и пр. В первом столбце таблицы записываются показания датчика пути (расстояние вдоль сварного шва в миллиметрах), а в остальных значения амплитуд сигналов по всем 16 схемам прозвучивания. Длина окружности трубопровода составляет 1020 мм. Для надежного контроля начального участка сканирование проводится с нахлестом от 10 до 100 мм. Амплитуда сигнала изменяется в диапазоне 0–255 условных единиц.

Пример файла с данными

Файлы с данными носят числовые названия, которые соответствуют номеру сварного шва при проведении УЗК. Форматом файлов с данными является .dat. Ниже показан пример одного такого файла.

```
0000 08-10-02 16:24:04 +33C   246   325   16   9 1 7У1      ЗОТ П2
0001 000 000 001 027 191 055 010 000 094 215 122 073 217 255 113 058
0002 000 015 008 031 134 090 009 002 116 222 080 146 179 255 105 080
0003 000 029 039 009 135 115 009 021 185 157 076 182 101 255 107 073
0004 003 029 109 000 255 164 017 033 255 121 098 186 112 255 118 095
0005 007 016 135 000 255 152 023 035 255 115 110 153 119 252 135 136
0006 016 003 087 000 255 118 025 031 255 078 127 114 119 234 139 089
0007 016 015 064 000 255 078 018 036 231 077 125 147 111 199 124 097
0008 009 028 079 000 244 065 034 045 255 096 109 186 098 224 158 110
0009 000 043 069 000 251 066 024 075 238 108 087 223 088 255 135 157
0010 001 064 041 006 255 061 000 073 178 147 094 255 067 255 079 150
0011 003 078 036 000 239 044 000 033 176 081 099 255 081 250 053 126
0012 005 088 046 000 212 037 000 054 200 027 102 255 141 255 042 149
0013 005 079 048 000 163 030 000 047 201 037 107 255 135 255 070 136
0014 015 069 032 000 160 058 001 040 162 041 129 255 113 255 089 163
0015 021 053 018 000 127 065 000 028 129 021 137 239 123 255 074 161
0016 021 045 015 000 072 028 015 042 126 000 139 225 195 242 121 146
0017 013 022 057 000 138 033 009 058 225 000 126 161 250 200 108 169
0018 008 005 096 000 159 000 034 082 255 000 108 119 255 153 155 170
0019 005 000 094 000 082 000 046 077 255 063 103 075 244 105 180 158
0020 003 000 067 000 131 007 031 042 234 005 098 036 222 066 147 150
0021 000 000 050 000 068 028 003 057 200 000 084 034 220 061 091 154
0022 000 000 069 000 041 011 000 041 242 000 063 043 214 114 068 131
0023 000 000 061 000 032 000 000 027 223 000 030 055 181 168 061 130
```

Рисунок 5 - Пример файла с данными

5 Целевая переменная

Целевую переменную нужно будет сделать из таблицы (csv формат), в столбцах которой будут указаны характеристики найденных дефектов.

Таблица дефектов

Начало дефекта	Длина дефекта	Высота дефекта	Тип дефекта	Сторона
20	30	4	L	B
130	35	5	L	A
256	29	3	L	A

310	38	7	L	B
515	25	6	L	A
830		7	T	
910		5	T	
178		3	T	

Комментарий к таблице:

Тип дефекта: L - протяженный, T - поперечный.

6 Обнаружение дефектов

Проявление дефекта в сигнале эхо-такта можно представить как увеличение уровня сигнала от некоторого начального значения, области постоянного уровня (при сканировании вдоль дефектной области) и последующим снижении уровня сигнала.

Длину дефекта определяют как разность координат конца и начала сигнала от дефекта на С-скане, то есть разность границ дефекта.

Высоту дефекта определяют по уровню падения сигнала от несплошности. Так например, если падение сигнала от 255 усл.ед. составляет 200 усл.ед. то высота дефекта находится в диапазоне от 8 мм и более. А если падение сигнала от 255 усл.ед. находится в диапазоне от 10 до 20 усл.ед. то высота дефекта будет 2-3 мм.

7 Тренировочная выборка

В качестве тренировочной последовательности будет использоваться выборка SOP, полученная в результате сканирования системой ПУЗК

стандартного образца предприятия (СОП). Сканирование образца выполнялось 3 раза подряд, именно поэтому даются 3 выборки для обучения.

Выполнение лабораторной работы

1. Прочтите данные из файлов в папке train и test. Файлы Sop1, Sop2 и Sop3 являются массивами данных обучающей выборки, а target1, target2, target3 – целевыми переменными для каждого файла Sop соответственно.

Функции, которые могут пригодиться при решении: pd.read_csv()

2. Отобразите несколько первых и несколько последних записей.

Функции, которые могут пригодиться при решении: .head(), .tail().

3. С помощью массивов, содержащих значения целевой переменной, создайте еще один вектор, содержащий названия сторон дефектов в местах их наличия. На всех остальных интервалах поставьте нулевые значения.
4. Разбейте данные из папки train на обучающую и проверочную (валидационную) выборки в пропорции 70 на 30 с помощью функции *train_test_split()* из библиотеки *sklearn*.
5. На тренировочной выборке обучите модель классификатора Байеса (GaussianNB) с параметрами, установленными по-умолчанию. Модель классификатора можно загрузить используя модуль библиотеки *sklearn – naive_bayes* (*from sklearn.naive_bayes import GaussianNB).*
6. Выполните предсказание на тестовой выборке из папки test. Оцените качество работы моделей с помощью метрики *accuracy* и *classification report* из библиотеки *sklearn* модуля *metrics.*
7. Выполните подбор гиперпараметров для модели классификатора с помощью *GridSearchCV() (from sklearn.model_selection import*

GridSearchCV) с параметром кросс-валидации cv = 5. Подумайте какие параметры стоит настроить. Аргументируйте свой выбор.

8. Заново обучите модель с подобранными гиперпараметрами на обучающей выборке и оцените качество ее работы на тестовой.
9. Оформите отчет по лабораторной работе в формате ipynb с заголовками, комментариями, рисунками (с заголовками и названиями осей), ответами на контрольные вопросы, а также выводами о проделанной работе. Перед первым заголовком должно быть ваше ФИО и название группы. Назовите файл ФИО_lab8.ipynb и сделайте файл .pdf с таким же названием, а затем сдайте оба файла преподавателю.

Контрольные вопросы

1. Что такое «Наивный байесовский классификатор? Опишите логику и этапы построения модели по алгоритму.
2. Почему байесовский классификатор называется «наивным»?
3. Что такое априорная и апостериорная вероятности?